BEST AVAILABLE IMAGES

Defective images within this document are accurate representation of The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68 — A1

(11) International Publication Number: WO 97/32996

(43) International Publication Date: 12 September 1997 (12.09.97)

(21) International Application Number: PCT/EP97/01070

(22) International Filing Date: 4 March 1997 (04.03.97)

(30) Priority Data:
08/611,587 6 March 1996 (06.03.96) US

(71) Applicants: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM) [FR/FR]; 101, rue de Tolbiac, F-75654 Paris Cédex 13 (FR). BAYLOR COLLEGE OF MEDICINE [US/US]; One Baylor Plaza, Houston, TX 77030-3948 (US).

(72) Inventors: PANDOLFO, Massimo; Apartment 112, 2111 Holly Street, Houston, TX 77054 (US). MONTERMINI, Laura; Apartment 1030, 2300 Old Spanish Trail, Houston, TX 77054 (US). MOLTO, Maria; Calle Corbella, 33.1, E-46240 Carlet (ES). KOENIG, Michael; 11, rue Aristide-Briand, F-67115 Plobsheim (FR). CAMPUZANO, Victoria; 7, rue de Berne, F-67000 Strasbourg (FR). COSSEE, Mireille; 5, rue Kirschleger, F-67000 Strasbourg (FR).

(74) Agent: PHELIP, Bruno; Cabinet Harlé & Phélip, 21, rue de la Rochefoucauld, F-75009 Paris (FR).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DIRECT MOLECULAR DIAGNOSIS OF FRIEDREICH ATAXIA

(57) Abstract

The invention relates generally to methods for the diagnosis and therapeutic treatment of Friedreich Ataxia. Friedreich ataxia (FRDA) is an autosomal recessive, degenerative disease that involves the central and peripheral nervous system and the heart. A gene, X25, was identified in the critical region for the FRDA locus on chromosome 9q13. The gene encodes a 210 amino acid protein, frataxin, that has homologues in distant species such as *C. elegans* and yeast. A few FRDA patients have been found to have point mutations in X25, but the vast majority are homozygous for a variable, unstable GAA trinucleotide expansion in the first X25 intron. Mature X25 mRNA was severely reduced in abundance in individuals with FRDA. Carriers and individuals at risk for developing FRDA can be ascertained by the methods of the present invention. Further, the methods of the present invention provide treatment to those individuals having FRDA.

AL

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

DIRECT MOLECULAR DIAGNOSIS OF FRIEDREICH ATAXIA

This invention was supported in part by a grant from the United States Government through federal funds (NINDS NS34192). The U.S. government has certain rights to this invention.

FIELD OF THE INVENTION

This invention relates generally to methods for the diagnosis, screening and therapeutic treatment of Friedreich ataxia. Friedreich ataxia (FRDA) is an autosomal recessive, degenerative disease that involves the central and peripheral nervous system and the heart. A gene, X25, was identified in the critical region for the FRDA locus on chromosome 9q13. The X25 gene encodes a 210 amino acid protein, frataxin, that has homologues in distant species such as *C. elegans* and yeast. A few FRDA patients have been found to have point mutations in X25, but the vast majority are homozygous for a variable, unstable GAA trinucleotide expansion in the first X25 intron. Mature X25 mRNA was severely reduced in abundance in individuals with FRDA.

BACKGROUND OF THE INVENTION

Friedreich ataxia (FRDA) is the most common hereditary ataxia, with an estimated prevalence of 1 in 50,000 and a deduced carrier frequency of 1/120 in the European population. FRDA is an autosomal recessive degenerative disease characterized by progressive gait and limb ataxia, a lack of tendon reflexes in the legs, loss of position sense, dysarthria, and pyramidal weakness of the legs. Hypertrophic cardiomyopathy is found in almost all patients. Diabetes mellitus is seen in about 10% of the cases. carbohydrate intolerance in an additional 20%. and a reduced insulin response to arginine stimulation in all cases. The age of onset is usually around puberty. and almost always before age twenty-five. Most patients are wheelchair bound by their late twenties and currently there is no treatment to slow progression of the disease.

The first pathologic changes are thought to occur in the dorsal root ganglia with loss of large sensory neurons, followed by deterioration of the sensory posterior columns, spinocerebellar tracts and corticospinal motor tracts of the spinal cord, and atrophy of large sensory fibers in peripheral nerves. Only occasional mild degenerative changes are seen in the cerebellum, pons and medulla. While most symptoms are a consequence of neuronal degeneration, cardiomyopathy and diabetes are thought to reflect independent sites of primary degeneration. Overall. the pathology of FRDA is very different from that of other hereditary ataxias. particularly the dominant forms and ataxia-telangiectasia. where the cerebellum is the primary site of degeneration.

The mutated gene in FRDA has been mapped to chromosome 9q13-q21.1. S. Chamberlain. et al., *Nature*, 334:248 (1988); and the FRDA candidate region has been narrowed to a 150 kb segment flanked by the Z0-2 gene (distal) and the marker F8101 (proximal), L. Montermini et al.,

Am. J. Hum. Genet., 57:1061 (1995). Previously proposed candidate genes are excluded: the X104/CSFA1/Z0-2 gene on the basis of the absence of deleterious mutation in patients, and the STM7 and PRKACG genes because they lie in entirety on the centromeric side of F8101 (Figure 1A).

SUMMARY OF THE INVENTION

It is a particular object of the present invention to provide a method of screening individuals for a mutation that leads to Friedreich's ataxia, comprising determining the number of GAA repeats in an intron of the X25 gene.

It is a further object of the present invention to provide a method of screening individuals for a mutation that leads to Friedreich's ataxia, comprising the steps of measuring expression of the X25 gene at the mRNA or protein levels.

It is another object of the present invention to provide a method of screening individuals for a mutation that leads to Friedreich's ataxia, comprising the step of detecting a variation in a size of a $(GAA)_n$ repeat in a first intron of a X25 gene by measuring the length of said repeat, wherein n for normal individuals ranges from 1-22 and n for affected individuals is more than about 120-900.

It is another object of the present invention to provide a method of screening individuals for a mutation that leads to Friedreich's ataxia, comprising the steps of sequencing DNA from an individual, and comparing said sequence from said individual to SEQ ID NOS 1-12 to determine what differences, if any, there are between the two sequences.

It is yet a further object of the present invention to provide a method of treating Friedreich's ataxia in an individual, comprising the step of

administering an effective pharmacologic dose of a protein having an amino acid sequence substantially similar to SEQ ID NO 4 to said individual.

It is an additional object of the present invention to provide a method of treating Friedreich's ataxia in an individual, comprising administration of a nucleic acid vector containing an X25 gene capable of expression in a pharmacologically acceptable carrier to said individual.

It is a further object of the present invention to provide compositions of matter having SEQ ID NOS 1-32.

Other and further objects. features and advantages will be apparent and the invention more readily understood from a reading of the following specification and by reference to the accompanying drawings forming a part thereof. wherein the examples of the presently preferred embodiments of the invention are given for the purposes of disclosure.

DESCRIPTION OF THE DRAWINGS

Figure 1(A): Transcription map of the FRDA critical interval. Distances are in kilobase pairs from the first Not I site upstream to the Z0-2 gene. The critical FRDA region is between the F8101 marker and the Z0-2 gene. M. Mlu I site; N, Not I site; E, Eag I site; S. Sac II site; B, BssH I site. Figure 1(B): Alignment of the exon 5a-containing isoform of frataxin with translated ORFs contained within a *C. elegans* cosmid (CELT59G1) and a *S. cerevisiae* EST (T38910). Identical amino acids are boxed. The putative signal peptide is underlined. Amino acids involved by point mutation (L106X and I154F) are indicated by vertical arrows. The exon 5b-containing isoform diverges at position 161, and its 11 COOH-terminal amino acids are RLTWLLWLFHP.

Figure 2: Northern blot analysis of X25 transcripts. A ^{32}P-labeled 5'-RACE product containing exons 1-5b was hybridized to a multiple tissue

Northern blot (Clontech), containing 2 µg of poly-A + RNA in each lane. The membrane was washed at 50° with 0.1 x SSC, 0.1% SDS, then exposed to x-ray film at -70° for 7 days. The lower panel shows a successive hybridization of the same blot with an actin probe (provided by the blot manufacturer).

Figure 3: Southern blot analysis showing FRDA-associated expanded restriction fragments. Lanes 1 and 12, normal controls; lanes 2-7, individuals from a Saudi Arabian FRDA family; lanes 8-11, individuals from a Louisiana Acadian (Cajun) FRDA family. Affected subjects are in lanes 3-5 and 9-10, heterozygous carriers in lanes 2, 6-8, and 11. The position of molecular weight markers is indicated on the side. The constant bands correspond to exons 2 and 3 (15 kb), and to a related sequence outside of the FRDA region (5 kb). Ten µg of genomic DNA from each individual were digested with Eco RI, run in a 0.6% agarose gel, and blotted onto a nylon membrane (Hybond+). The blot was hybridized with a ^{32}P-labeled X25 cDNA probe. After a highest stringency wash with 0.1 x SSC, 0.1% SDS for 5' at 65°, the blot was exposed to x-ray film at -70° for two days.

Figure 4: An automated sequence of the FRDA-associated expanded region from a cosmid subclone. The CTT strand was sequenced.

Figure 5: Automated sequence of the FRDA-associated expanded region containing the expanded repeat in a FA patient. The CTT strand was sequenced. It is interesting to note the presence of two imperfect repeats in the patient (the 7th and 8th in the sequenced strand) that are not present on the normal sequence and which could indicate a polymorphic variant present on the chromosome in which the original expansion occurred.

Figure 6(A): Example of PCR analysis of normal alleles of the GAA repeat. Lane 1 is the 1kb ladder DAN size marker, lanes 2-6 are normal controls previously identified to be heterozygous at the repeat. The GAA-F/GAA-R primers were used for amplification. Fragments vary in size in the 480-520 bp range.

Figure 6(B): PCR amplification of the expanded GAA repeat in a FRDA carrier (lane 3) and in a patient (lane 4). Lane 1 is the 1 kb ladder DNA marker, lane 2 is a normal control. The Bam/2500 primers were used for PCR. Expanded alleles have a slightly fuzzy appearance. Instability of the repeat is indicated by the presence of two distinct bands in the patient lane, although the patient is an offspring of consanguineous parents. Also, the carrier in lane 3 is the patient's mother, but the corresponding expanded allele does not exactly match in size any of her offspring bands.

Figure 7: Segregation of the L106X mutation and of the GAA expansion in a FRDA family. The SSCP pattern shown in A indicates the paternal origin of the point mutation, while Southern blot analysis, shown in B, indicates the maternal origin of the expansion. NR indicates an unrelated normal control.

Figure 8 : RT-PCR analysis of X25 mRNA in FRDA subjects, obligate carriers and normal controls. Reactions were performed on total RNA extracted from lymphoblastoid cell lines. The serine hydroxymethyltransferase (SHMT) transcript (encoded by a gene on chromosome 17) was used as a control for RNA amount. Mock reactions without reverse transcriptase (-RT) were also performed as a negative control. In the case of SHMT, the PCR following the -RT reactions generated a product of larger size than the product expected from the cDNA because a fragment of genomic DNA (contaminating the RNA preparation) containing a small intron was amplified. In all three

panels the lane marked with r.t. is a negative control (water), lane 9 corresponds to a normal control individual, lanes 1 and 4 to obligate carriers of FRDA, lanes 2, 3, and 5 to 8 to individuals with FRDA. To generate cDNA from the X25 transcript, the RT reaction was primed with the oligonucleotide E2R (SEQ ID NO 13), then PCR was performed between this primer and the nF primer (SEQ ID NO 14).

DETAILED DESCRIPTION OF THE INVENTION

It will be apparent to one skilled in the art that various substitutions and modifications may be made to the invention disclosed herein without departing from the scope of the invention.

As used herein, "FRDA" refers to Friedreich ataxia, an autosomal recessive, degenerative disease that involves the central and peripheral nervous system as well as the heart.

As used herein, "GAA expansion" refers to multiple $(GAA)_n$ repeats located 1.4 kb downstream from exon 1 in an intron of the X25 gene.

As used herein, the "X25" gene refers to the gene identified on chromosome 9q13 that is in the critical region of the FRDA-determinative locus .

As used herein the term "polymerase chain reaction" or "PCR" refers to the PCR procedure described in the patents to Mullis, et al., U.S. Patent Nos. 4,683,195 and 4,683,202. The procedure basically involves: (1) treating extracted DNA to form single-stranded complementary strands; (2) adding a pair of oligonucleotide primers, wherein one primer of the pair is substantially complementary to part of the sequence in the sense strand and the other primer of each pair is substantially complementary to a different part of the same sequence in the complementary antisense strand; (3) annealing the paired primers to the complementary sequence; (4) simultaneously extending the annealed primers from a 3` terminus of each primer to synthesize an

extension product complementary to the strands annealed to each primer wherein said extension products after separation from the complement serve as templates for the synthesis of an extension product for the other primer of each pair; (5) separating said extension products from said templates to produce single-stranded molecules: and (6) amplifying said single-stranded molecules by repeating at least once said annealing, extending and separating steps.

As used herein. the term "pulsed field gel electrophoresis" or "PFGE" refers to a procedure described by Schwartz, et al., *Cold Springs Harbor Symposium, Quantitative Biology*, 47:189-195 (1982). The procedure basically comprises running a standard electrophoresis gel (agarose, polyacrylamide or other gel known to those skilled in the art) under pulsing conditions. One skilled in the art recognizes that the strength of the field as well as the direction of the field is pulsed and rotated in order to separate megabase DNA molecules. Current commercial systems are computer controlled and select the strength, direction and time of pulse depending on the molecular weight of DNA to be separated.

As used herein. the phrase "gene transcript" shall mean the RNA product that results from transcribing a genetic (DNA) template. "Gene" shall mean a hereditary unit: in molecular terms, a sequence of chromosomal DNA that is required for production of a functional product.

As used herein. the phrase "messenger RNA" or "mRNA" shall mean an RNA transcribed from the DNA of a gene, that directs the sequence of amino acids of the encoded polypeptide.

As used herein. the phrase "copy DNA" or "cDNA" shall mean DNA synthesized from a primer hybridized to a messenger RNA template.

As used herein, the phrase "oligonucleotide" shall mean a short nucleic acid molecule (usually 8 to 50 base paris), synthesized for use as a probe or primer.

As used herein, the phrase "primer" shall mean a short DNA or RNA molecule that is paired with a complementary DNA or RNA template, wherein the short DNA or RNA molecule provides a free 3'-OH terminus at which a DNA polymerase starts synthesis of a nucleotide chain.

It is a particular object of the present invention to provide a method of screening individuals for a mutation that leads to Friedreich's ataxia, comprising the steps of digesting DNA from an individual to be tested with a restriction endonuclease; and measuring the length of a restriction fragment length polymorphism (RFLP) by hybridization to probes that recognize a region encompassing a GAA repeat in a first intron of an X25 gene and performing Southern Blot analysis, wherein an RFLP corresponding to a GAA repeat longer than a normal range of 7-22 triplets, usually more than about 120, is an indication of said mutation that leads to Friedreich's ataxia.

It is a further object of the present invention to provide a method of screening individuals for a mutation that leads to Friedreich's ataxia, comprising the steps of measuring expression of an X25 gene by determining an amount of mRNA expressed from the X25 gene and from known controls, and comparing the amount of mRNA from the X25 gene to the amount of mRNA from the known controls, wherein a reduced amount of mRNA from the X25 gene indicates individuals having said mutation that leads to Friedreich's ataxia.

It is an additional object of the present invention to provide a method of screening individuals for a mutation that leads to Friedreich's ataxia, wherein the amounts of mRNA is determined by the steps of extracting mRNA from individuals to be tested; preparing cDNA from said mRNA, amplifying said cDNA to produce amplification products; and comparing relative amounts of X25 and control cDNA present, wherein a reduced amount of mRNA from the X25 gene indicates individuals having said mutation that leads to Friedreich's ataxia.

It is an additional object of the present invention to provide a method of screening individuals for a mutation that leads to Friedreich's ataxia. by detecting the amount of specific proteins encoded by X25 in cells from patients using antibodies specific for the X25 proteins.

It is another object of the present invention to provide a method of screening individuals for a mutation that leads to Friedreich's ataxia, comprising the step of detecting a variation in a size of a $(GAA)_n$ repeat in a first intron of a X25 gene by measuring the length of said repeat, wherein n for normal individuals ranges from 1-22 and n for affected individuals is more than about 120-900.

It is an additional object of the present invention to provide a method for detecting a GAA polymorphism in a first intron of an X25 gene comprising the steps of performing a PCR assay to produce amplified products of said first intron of said X25 gene and measuring the length of said amplified products with molecular techniques known in the art.

It is another object of the present invention to provide a method of screening individuals for a mutation that leads to Friedreich's ataxia, comprising the steps of sequencing DNA from an individual, and comparing said sequence from said individual to SEQ ID NOS 1-12 to determine what differences. if any, there are between said sequence from said individual and said SEQ ID NOS 1-12.

It is yet a further object of the present invention to provide a method of treating Friedreich's ataxia in an individual, comprising the step of administering a pharmacologic effective dose of a protein having an amino acid sequence substantially similar to SEQ ID NO 4 to said individual.

It is an additional object of the present invention to provide a method of treating Friedreich's ataxia in an individual, comprising administration of a nucleic acid vector containing an X25 gene capable of expression and a pharmacologically acceptable carrier to said individual.

It is a further object of the present invention to provide compositions of matter having SEQ ID NOS 1-32.

The therapeutic compositions of the present invention can be formulated according to known methods to prepare pharmacologically useful compositions. The compositions of the present invention or their functional derivatives are combined in admixture with a pharmacologically acceptable carrier vehicle. Suitable vehicles and their formulations are well known in the art. In order to form a pharmacologically acceptable composition suitable for effective therapeutic administration, such compositions will contain an effective amount of the X25 gene or its equivalent or the functional derivative thereof, or the frataxin protein or its equivalent or the functional derivative thereof, together with the suitable amount of carrier vehicle.

The nucleic acid therapeutic composition of the present invention will usually be formulated in a vector. The frataxin protein therapeutic composition will usually be administered as a purified protein in a pharmacologically suitable carrier.The compositions can be administered by a variety of methods including parenterally, by injection, rapid infusion, nasopharyngeal absorption, dermal absorption or orally. The compositions may alternatively be administered intramuscularly or intravenously. In addition, the compositions for parenteral administration can further include sterile aqueous or nonaqueous solutions, suspensions and emulsions. Examples of known nonaqueous solvents include propylene glycol, polyethylene glycol, vegetable oils, such as olive oil, and injectable organic esters, such as ethyl oleate. Carriers, adjuncts or occlusive dressings can be used to increase tissue permeability and enhance absorption. Liquid dosage forms for oral administration may generally comprise a liposome solution. Suitable forms for suspension include emulsions, solutions, syrups and elixirs containing inert diluents commonly used in the art, such as purified

water. Besides the inert dilutants, such compositions can also include wetting agents, emulsifying and suspending agents or sweetening, flavoring, coloring or perfuming agents.

Additionally, pharmaceutical methods may be employed to control the duration of action. These are well known in the art and include control release preparations and can include appropriate macromolecules, for example polymers, polyesters, polyamino acids, polyvinyl, pyrolidone, ethylenevinylactate, methylcellulose,carboxymethylcellulose,or protamine sulfate. The concentration of macromolecules, as well as the methods of incorporation, can be adjusted in order to control release. Additionally, the vector could be incorporated into particles of polymeric materials such as polyesters, polyamino acids, hydrogells, poly (lactic acid) or ethylene vinylacetate copolymers. In addition to being incorporated, these agents can also be used to trap the vectors in microcapsules. These techniques are well known in the art.

A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in detectable change in the physiology of a recipient patient.

Generally, the dosage needed to provide an effective amount of composition will vary depending on such factors as the recipient's age, condition, sex and extent of disease, if any, and other variables which can be adjusted by one of ordinary skill in the art.

One skilled in the art will appreciate readily that the present invention is well adapted to carrying out the ends and advantages mentioned as well as those inherent herein. The probes, primers, methods, procedures and techniques described are presently representative of the preferred embodiments, are intended to be exemplary, and are not

intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art, and are encompassed within the spirit of the invention or defined by the scope of the appended claims. All references specifically cited herein are incorporated by reference.

The following examples are offered by way of illustration and are not intended to limit the invention in any manner:

Example 1: Localizing and Sequencing the FRDA Critical Region

Potential exons were identified in the FRDA critical region by direct cDNA selection, exon amplification, and computer prediction from random sequences. Twelve cosmids spanning 120 kb of the critical FRDA interval, plus 80 kb immediately proximal to the interval were subcloned individually as Bam HI - Bgl II fragments into pSPL1 and pSPL3 exon-trapping vectors and transfected into COS-7 (A6) cells for splicing of potential exons. See D.M. Church et al., *Nature Genet.* 6:98 (1994). The same cosmids were used for hybridization-selection from uncloned cDNAs synthesized from human cerebellum and spinal cord poly-A+RNA. See J.G. Morgan et al., *Nucl. Acids Res.* 20:5173 (1992). Finally, seven of the cosmids were subcloned as Sau 3AI, Apo I, and Hae III fragments and about 1500 random single pass sequences were generated. These sequences were analyzed using the GRAIL1a and GRAIL2 (E.C. Uberbacher and R.J. Mural, *Proc. Natl. Acad. Sci. USA* 88:11261 (1991). and FEXH (V.V. Solovyev et al., *Nucl. Acids Res.* 22, 5156 (1994)) programs.

These analyses yielded 19, 5, and 17 potential coding sequences, respectively, including two that matched known genes, namely the protein kinase A gamma catalytic subunit gene (obtained by cDNA selection and random sequencing), and a mitochondrial adenylate

kinase 3 pseudogene (obtained by random sequencing). One exon, called d26, was identified independently by two approaches. Nested primers based on the d26 sequence, when used in a rapid amplification of cDNA 5' end (5'-RACE) experiment on a heart cDNA template yielded two independent but overlapping products. The 5'-RACE was performed using the Clontech RACE-ready cDNA kit according to the manufacturer's instructions. Sequence from these clones matched another amplified exon and an expressed sequence tag (EST) from a human liver+spleen cDNA library (Homosapiens cDNA clone 126314, 5' sequence (GenBank accession number R06470)). This gene, called X25, apparently had alternate transcripts, because the sequences at the 3' end of the EST and RACE products were different.

The gene structure of X25 (Figure 1A) was resolved by obtaining intronic sequences flanking the identified exons, by inverse PCR, and by direct sequencing of cosmids. The EST clone contained 4 exons, and the longer RACE product contained one additional 5' exon. This exon mapped with the CpG island at position 100 on the genomic map. A transcription start site was predicted 388 bp upstream of the exon I donor splice site, and a TATA box was found 28 bp further upstream by the TSSG program. Five exons (1 to 5a, where exon 5a corresponds to the 3' end of the EST) were found to be spread over 40 kb. They contain an open reading frame (ORF) encoding a 210 amino-acid protein, which was named frataxin (Figure 1B). An alternative exon (5b), corresponding to d26, was localized at about 40 kb from exon 5a in the telomeric direction. Exon 5b also has an in-frame stop codon, so that the alternative transcript encodes a shorter, 171 amino acid protein, whose 11 COOH- terminal residues differ from the main isoform. Nucleotide sequences of the X25 exons have been deposited in the GenBank database under the accession numbers U43748 to U43753. The 3' end of the transcript encoding the alternative form was

investigated by 3' RACE (see Froman and Martin. *Technique* 1:165 (1989)), 2 µg total RNA from Hela cells was used with nested primers in exon $5b_3$ and showed that, depending on the alternate usage of the 3' donor splice site in exon 5b, either a transcript ending with this exon, or a longer transcript including an additional non-coding exon 6 could be generated. This longer 3' RACE product ended with the poly-A tail of a downstream Alu sequence. Genomic sequence of exon 6 showed that it contains 3 Alu sequences in tandem, followed by a polyadenylation signal 1050 bp away form the acceptor splice site. Exon 6 was mapped 13 kb telomeric to exon 5b (Fig. 1A). Splice sites of all 7 exons (1 to 4, 5a, 5b, and 6) conform to the canonical consensus.

Example 2: Expression of the X25 Transcript.

Poly A+ Northern blots of different human tissues revealed the highest expression of X25 in heart, intermediate levels in liver, skeletal muscle, and pancreas, and minimal expression in other tissues, including whole brain (Figure 2). A 1.3 kb major transcript was identified, in agreement with the predicted size of an exon 5a-containing mRNA. Fainter bands of 1.05, 2.0, 2.8, and 7.3 kb were also detected. Further hybridizations of the northern blot with exon 5a- and 5b-specific probes revealed that the 1.05 and 2.0 kb bands contained exon 5b, while sequences matching exon 5a were found in the 1.8 and 7.3 kb bands in addition to the major 1.3 kb band. A northern blot of total RNA from selected parts of the central nervous system (CNS) revealed high expression of the 1.3 kb transcript in the spinal cord, with less expression in cerebellum, and very little in cerebral cortex (not shown). Overall, expression of X25 appeared to be highest in the primary sites of degeneration in FRDA, both within and outside the CNS.

To investigate the nature of the larger transcripts, a fetal brain cDNA library was screened with the EST clone (exons 2-5a). Among nine positives, four clones were isolated whose sequence extended beyond the limits of the previously identified X25 mRNAs. Sequence analysis of these clones indicated that they originated from a related gene, differing from X25 at several positions, and with stop codons in the sequence corresponding to X25 exon 1. Three of the cDNAs, which are identical in the portion that has been sequenced, extend respectively for 0.5, 1 and 2 kb upstream of exon 1. Their sequence presents numerous divergences from X25 in the part corresponding to exon 1, mostly CpG dinucleotides changed in TG or CA, then being almost identical in the part corresponding to exons 2 to 4.

An additional 1.6 kb cDNA begins with a sequence closely matching exon 5a, even in its UTR, with only occasional single base changes and short insertions/deletions. The X25 related gene was excluded from the critical FRDA region, and at least one intronless copy exists in the genome, as indicated by Southern blot and PCR analysis. Southern blot analysis with an X25 exon 1-5a cDNA probe revealed a prominent 5 kb Eco RI band in genomic DNA that did not correspond to any exon and was absent in YAC and cosmid DNA from the critical FRDA region. Several additional bands, also absent from cloned DNA from the FRDA region, appeared when blots were washed at lower stringency (1 X SSC at room temperature). The primers nF2 (5'-TCCCGCGGCCGGCAGAGTT-3') [SEQ ID NO 14] and E2R (5'-CCAAAGTTCCAGATTTCCTCA-3') [SEQ ID NO 13], which can amplify a 173 bp fragment spanning exons 1 and 2 of the X25 cDNA, generated a PCR product of corresponding size from genomic DNA, but not from cloned DNA from FRDA region, indicating the presence of sequences with high similarity to a processed X25 transcript elsewhere in the genome.

Example 3: Computer Database Search.

A BLASTN DNA database search with the X25 DNA sequence and a BLASTP search with the translated sequence did not reveal any significant match. However, a TBLASTN search in which the protein sequence was compared to the six-frame translation of the DNA databases yielded highly significant matches with an ORF contained in a *C. elegans* cosmid ($P = 7.6 \times 10^{-13}$) and with a *S. cerevisiae* EST ($P=2.0 \times 10^{-10}$) (Fig. 1B). In both cases, the closest match involved a 27-aa segment of the protein (positions 141-167) encoded in exons 4 and 5a, showing 25/28 and 22/27 amin-oacid identity with the *C. elegans* and *S. cerevisiae* sequences, respectively, and 65% identity at the DNA level. Secondary structure predictions for the X25-encoded protein suggested an α-helical structure for the NH_2-terminal 30 amino acids and the regions between residues 90-110 and 185-195, with possible interspersed β-sheet regions around residues 125-145 and 175-180. Secondary structure prediction was performed with the SSP and NNSSP programs. which are designed to locate secondary structure elements (V.V. Solovyev and A.A. Salamov. *CABIOS* 10:661 (1994)). The TMpred program was used to predict putative transmembrane domains (K. Hoffmann and W. Stoffel. *Biol. Chem. Hoppe-Seyler* 374:166 (1993)). PSORT was used to predict possible protein sorting signals (K. Nakai and M. Kanehisa. *Proteins: Structure, Function, and Genetics* 11:95 (1991)). No transmembrane domain was identified. As computer analysis of the amino acid sequence suggests that the frataxin protein contains an N-terminal hydrophobic signal, it may be a precursor for a secreted protein with a growth factor or hormone-like action, making frataxin an ideal protein for expression in bacteria, yeast and mammalian cells.

Example 4: Determining the Nature of the Mutation Leading to FRDA.

All six coding exons of X25 in 184 FRDA patients were amplified with flanking primers and screened for mutations. The following intronic primers were used to amplify the X25 exons: exon 1 (240 bp), F: 5'-AGCACCCAGCGCTGGAGG-3'[SEQ ID NO15], R: 5'-CCGCGGCTGTTCCCGG-3' [SEQ ID NO 16]; exon 2 (168 bp), F: 5'-AGTAACGTACTTCTTAACTTTGGC-3' [SEQ ID NO 17]; R: 5'-AGAGGAAGATACCTATCACGTG'-3' [SEQ ID NO 18], exon 3 (227 bp), F: 5'- AAAATGGAAGCATTTGGTAATCA-3' [SEQ ID NO 19], R: 5'-AGTGAACTAAAATTCTTAGAGGG-3' [SEQ ID NO 20]; exon 4 (250 bp), F: 5'-AAGCAATGATGACAAAGTGCTAAC-3' [SEQ ID NO 21]; R: 5'-TGGTCCACAATGTCACATTTCGG-3' [SEQ ID NO 22]; exon 5a (223 bp), F: 5'- CTGAAGGGCTGTGCTGTGGA-3' [SEQ ID NO 23], R: 5'-TGTCCTTACAAACGGGGCT-3' [SEQ ID NO 24], exon 5b (224 bp), F: 5'-CCCATGCTCAAGACATACTCC-3' [SEQ ID NO 25], R: 5'-ACAGTAAGGAAAAAACAAACAGCC-3' [SEQ ID NO 26]. Amplifications for exons 2. 3. 4. 5a. and 5b consisted of 30 cycles using the following conditions: 1 min. at 94°. 2 min. at 55°, 1 min. at 72°. To amplify the highly GC-rich exon 1, the annealing temperature was raised to 68° and 10% DMSO was added to the reaction. The search for mutations was conducted using single-strand conformation polymorphism (SSCP) analysis (see M. Orita et al., *Genomics* 5:874 (1989)) in 168 FRDA patients, and chemical clevage (see J. A. Saleeba et al., *Hum. Mutat.* 1:63 (1992)) in 16. Three point mutations that introduce changes in the X25 gene product were identified.

Point Mutations. The first change, in a French family with two affected siblings, consisted of a T→ G transversion in exon 3 that changed a leucine codon (TTA) into a stop codon (TGA)(L106X). The second case, in a Spanish family with one affected member, was an A→G transition that disrupted the acceptor splice site at the end of the third intron, changing the invariant AG into a GG. Finally, a change from isoleucine to phenylalanine (I154F) was found in exon 4 in five patients from three Southern Italian families. This conservative change of an hydrophobic amino acid affects an invariant position within the highly conserved domain shared between human, worm and yeast. In all three cases, affected individuals were heterozygous for the point mutation. The I154F mutation was also found in 1 out of 417 chromosomes from 210 control individuals from the same Southern Italian population, which is compatible with the possibility that this is a disease-causing mutation. (Assuming a FRDA carrier frequency in Italy of 1/120 individuals and a frequency of I154F of 1/40 FRDA chromosomes in Southern Italians, one individual in 3,300 in that population is expected to be a carrier of 1154F. Finding such an individual in a random sample of 210 subjects can occur with >6% probability.)

Intron 1 Expansion. Southern blot analysis did not reveal any difference between FRDA patients and normal controls, when DNAs digested with Msp I, Taq I, or Bst XI were hybridized with an X25 cDNA probe, thereby excluding major rearrangements. Hybridization of Eco RI-digested DNAs from FRDA patients, however, revealed that the fragment containing exon 1 was on average 2.5 kb larger than in normal controls, with no detectable normal band. FRDA carriers were heterozygous for an enlarged- and a normal-sized fragment. The size of the enlarged fragment was clearly variable, even among FRDA carriers who were related (Figure 3). The enlarged region was localized further

to a 5.2 kb Eco RI/Not I fragment within the first intron of X25. which was subcloned from a cosmid and sequenced.

Oligonucleotide primers were designed to amplify this fragment using a long-range PCR technique, and its increased in size in FRDA patients was confirmed. The Perkin-Elmer XL long-PCR reagent kit was used to set up the reactions, utilizing standard conditions as suggested by the manufacturer and primers 5200Eco (5'- GGGCTGGCAGATTCCTCCAG-3') [SEQ ID NO 27] and 5200Not (5'-GTAAGTATCCGCGCCGGGAAC-3') [SEQ ID NO 28]. Amplifications were performed in a Perkin-Elmer 9600 machine, and consisted of 20 cycles of the following steps: 94° for 20 sec., 68° for 8 min.. followed by further 17 cycles in which the length of the 68° increased by 15 sec./cycle. The generated amplification product is 5 kb from normal chromosomes, and about 7.5 kb from FRDA chromosomes.

Cosmid sequence analysis revealed a $(GAA)_9$ repeat apparently derived from a poly-A expansion of the canonical A_5TACA_6 sequence linking the two halves of an Alu repeat (Figure 4 showing the reverse complementary sequence). The $(GAA)_9$ repeat is located 1.4 kb downstream from exon 1, and restriction analysis of long-range PCR fragments from FRDA patients located the abnormal size increase within 100 bp from this triplet repeat. Digestion of the same fragments with Mbo II, whose recognition site is GAAGA, suppressed size difference between patients and controls, indicating that the GAA repeat may be involved. Direct sequencing proved that the mutation consists of an almost pure GAA repeat expansion (Figure 5). PCR primers were then designed to evaluate the presence and size of the GAA expanded repeat FRDA patients, and any variability of the repeat in normal individuals (Figure 6).

The primers GAA-F (5'-GGGATTGGTTGCCAGTGCTTAAAAGTTAG-3')[SEQ ID NO29] and GAA-R (5'-

ATCTAAGGACCATCATGGCCACACTTGCC-3') [SEQ ID NO 30] flank the GAA repeat and generate a PCR product of 457 + 3n bp (n = number of GAA triplets). With these primers, efficient amplification of normal alleles could be obtained by using the traditional PCR procedure with Taq polymerase, after 30 cycles consisting of the following steps: 94° for 45 sec., 68° for 30 sec., 72° for 2 min. Enlarged alleles were much less efficiently amplified, particularly when present together with a normal allele; therefore, use of these primers is not indicated for FRDA carrier detection. A more efficient amplification of expanded alleles, also in FRDA carriers, is obtained using the primers Bam (5'-GGAGGGATCCGTCTGGGCAAAGG-3') [SEQ IDNO 31] and 2500F (5'-CAATCCAGGACAGTCAGGGCTTT-3') [SEQ ID NO 32]. These primers generated a ~1.5 kb (1398 bp) normal fragment. Amplification was conducted using the long PCR protocol, in 20 cycles composed of the following steps: 94° for 20 sec., 68° for 2 min. and 30 sec., followed by further 17 cycles in which the length of the 68° step was increased by 15 sec/cycle.

Seventy-nine unrelated FRDA patients with typical disease, including five patients known to carry X25 point mutations, were tested for the GAA expansion by Southern analysis and/or by PCR. The patients previously known to carry point mutations were all heterozygous for the expansion. Segregation analysis within families indicated that the point mutation and the GAA expansion had different parental origin (Figure 7), demonstrating that the point mutations--including the conservative missense mutation I154F--are disease causing. Homozygosity for expanded alleles was demonstrated in 71 of the 74 patients without previously-detected X25 point mutations, and heterozygosity was demonstrated in three.

Overall, according to these data the GAA expansion accounted for about 98% of the FRDA phenotype. The sizes of the enlarged alleles

were found to vary between 200 and more than 900 GAA units, with most alleles containing 700-800 repeats. Instability of expanded repeats during parent-offspring transmission was clearly demonstrated; both directly by analysis of parent-offspring pairs. and indirectly by the detection of two distinct alleles in affected children of consanguineous parents, who are expected to be homozygous-by-descent at the FRDA locus. PCR products corresponding to expanded repeats appeared as slightly blurred bands. suggesting the occurrence of only a limited degree of somatic mosaicism for different size repeats due to mitotic instability, at least in lymphocyte DNA (Figure 6B). Seventy-seven normal individuals who were tested by Southern analysis were homozygous for a normal allele. PCR analysis of additional 98 normal controls also did not show any expansion, and revealed that the GAA repeat is polymorphic, its length varying from 7 to 22 units (Figure 6A). Smaller alleles were more prevalent.

GAA repeats, up to 30-40 units, are common in many organisms and are sometimes polymorphic, as in the 3' UTR of the rat polymeric Ig receptor; however. they have not previously been associated with disease. A recently proposed theoretical model suggests that ability to form a hairpin structure is crucial for the susceptibility of trinucleotide repeats to give rise to large expansions (See A. M. Gray et al., *Cell* 81:533 (1995)). According to this model, CAG/CTG or CGG/CCG repeats are predicted to be expansion prone, while the GAA/CCT repeat has lowest propensity to expand. making the FRDA expansion an unexpected finding. A striking linkage disequilibrium between FRDA and a polymorphism in a newly-identified exon of the Z0-2 gene (about 120 kb telomeric to the expanded triplet repeat) in French and Spanish families suggests a single origin for the FRDA expansion, but it is also compatible with a multistep or recurrent expansion on an allele at risk. (See Imbert et al.. *Nature Genet.* 4:72 (1993) where the absolute linkage

disequilibrium in myotonic dystrophy is expanded by recurrent mutations on such a risk allele.) The fact that RDA is autosomal recessive makes the natural history of the mutation at the population level strikingly different from any other known disease due to trinucleotide expansions.

In fragile X and myotonic dystrophy, where expansions of comparable size occur in non-coding sequences, carriers have severe early-onset disease and a strong reproductive disadvantage. Large expansions in these diseases are newly formed from unstable alleles of intermediate sizes, resulting in the phenomenon of anticipation. In FRDA, large expanded alleles are transmitted by asymptomatic carriers, and new expansion events in heterozygotes would go undetected at the phenotypic level. Absence of negative selection against heterozygotes plays the key role in maintaining the frequency of large FRDA expanded alleles as high as 1 per 250 chromosomes, at least one order of magnitude higher than any other characterized trinucleotide expansion.

Conversely, deletions of CTG repeats in myotonic dystrophy with reversion to normal size alleles have been observed (see Imbert et al., *Nature Genet.* 4:72 (1993)) wherein the absolute linkage disequilibrium in myotonic dystrophy is explained by recovered mutations in such a risk allele. In the sample of FRDA families in the study of the present invention, large expanded alleles were present in all tested symptomatic carriers, and, despite their size instability, neither new expansions deriving from an intermediate allele nor reversions to normality were detected. Although the occasional occurrence of such events cannot be excluded in the general population given the large number of heterozygous individuals, it appears that the frequency is low enough not to introduce detectable distortions in the pattern of FRDA inheritance, particularly inconsistencies in linkage results.

Example 5: Quantification of the FRDA Transcript.

When the X25 transcript was amplified with primers connecting exons 1 and 2, FRDA patients showed either undetectable or extremely low mRNA levels when compared to carriers and unrelated controls.

RT-PCR. RT-PCR was done on lymphoblast RNA from two normal controls, two obligate carriers, and six patients, using the exon 2 reverse primer E2R (5`- CCAAAGTTCCAGATTTCCTGA-3`) [SEQ ID NO 13] and the exon 1 forward primer nF (5`-CAGGCCAGACCCTCAC-3`) [SEQ ID NO 14]. As a precaution to avoid amplification of X25-related sequences not deriving from the FRDA region transcript, the nF primer was chosen to have no match with the non-9q13 related gene. PCR reactions were carried out for 25 cycles in order to maintain linearity between starting and final concentrations of DNA fragments. PCR products were blotted onto nylon membranes and hybridized with the ^{32}P end-labeled internal oligonucleotide nF2 ('5-TCCCGCGGCCGGCAGAGTT-3') [SEQ ID NO 15]. This observation suggests that either an abnormality in RNA processing, or an interference with the transcription machinery, occur as a consequence of the intronic GAA expansion.

Patients with deleterious point mutations affecting X25 clearly demonstrate that no other gene in the region, which could, in principle, be affected by a GAA expansion, is involved in the causation of FRDA. The restricted expression of X25 in the sites of degeneration or malfunction distinguishes FRDA from the dominant ataxias and from ataxia telangiectasia, where expression of the causative gene is ubiquitous. A severely reduced X25 mature mRNA is expected to result in a similarly low level of frataxin. Reduced frataxin in spinal cord, heart

and pancreas is likely the primary cause of neuronal degeneration, cardiomyopathy and increased risk of diabetes.

RNase Protection. In order to synthesize antisense riboprobes, two regions of the X25 cDNA were subcloned in a plasmid vector containing the T7 RNA polymerase promoter. Two separate segments of then X25 cDNA, one containing exons 1 and 2 (partial) and the other containing exons 4 (partial) and 5b were subcloned accordingly. 1 µg of linearized plasmid was used as a template for *in vitro* transcription (using the Ambion Maxiscript kit) in a reaction containing 3 µM α- ^{32}P UTP. The reaction was carried out at 37°C for an hour, after which the DNA template was completely digested by RNase-free DNase treatment. Full-length labeled transcripts were then purified following proparative denaturing polyacrylamide gel electrophoresis. A human GAPDH riborprobe (pTRI-GAPDH human, Ambion) was also generated as a control.

The RNase protection assay was performed using the RPAII Ribonuclease protection assay kit from Ambion following the manufacturer's recommendations. Briefly, 20 µg of total RNA extracted from patient and control lymphoblastoid cell lines was mixed with 8 x 10^4 cpm-labeled riboprobe in a 20 µl reaction, denatured and allowed to incubate at 45°C for 16 hours. 2 µg of RNA was used for the control GAPDH reaction. For each riboprobe, yeast RNA control hybridizations were performed as well. RNase (RNase A/RNase T1 mixture) treatment was carried out for 30 minutes at 37°C. The reaction products were ethanol precipitated and resuspended in formamide loading dye. These products were denatured and electrophoresed on a pre-heated 5% polyacrylamide/8 M urea gel in 1x Tris-borate buffer at 35 watts constant power.The gel was dried and exposed to an X-ray film for 6 days at -70°C using intensifying screens. The sizes of the protected fragments

were estimated accurately using a sequence ladder that had been co-electrophoresed with the sample.

Example 6: Therapeutics.

FRDA is caused by abnormalities in the X25 gene leading to a deficiency of its protein product, frataxin, occasionally due to point mutations that generate a truncated protein but, most commonly, to a GAA expansion in the first intron that causes supression of gene expression. Therapeutic administration of frataxin to FRDA patients is therefore an aspect of the present invention. Large amounts of recombinant frataxin is produced by cloning X25 cDNA into an expression vector that is tranformed into a suitable organism. Expression vectors that lead to production of high amounts of recombinant protein can be purified by several techniques. and prepared for systemic or local administration to patients. Computer analysis of the frataxin sequence suggests that frataxin is not a membrane protein, and is likely secreted. Both characteristics make frataxin an ideal protein for administration.

Another approach is examining the function of the frataxin protein and identifying compounds that can induce a cellular response or modification in the cell metabolism in cells that produce and/or respond to frataxin. Such compounds overcome the consequences of the lack of frataxin protein in FRDA.

Additionally, one can inactivate the murine X25 homolog via homologous recombination to provide an animal model for Friedreich's ataxia in order to test various therapeutic strategies.

Finally, the coding sequence for frataxin is inserted into a suitable expression vector that is administered to FRDA patients. The coding sequence of frataxin is inserted in the genome of a modified RNA or DNA virus, which is administered systemically or locally to patients, or

used to transduce cultured cells from patients that are then re-implanted into the patient body. Alternatively, non-viral vectors are utilized and administered directly to the patients or to patient's cultured cells that are re-implanted into the patient.

One skilled in the art will appreciate readily that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The nucleotides, proteins, peptides, methods, procedures and techniques described herein are presently representative of the preferred embodiments, are intended to be exemplary and are not intended as limitations on the scope. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention or defined by the scope of attached claims.

SEQUENCES

SEQUENCE ID (SEQ ID NO 1)

```
1    GATCGAGAAT AGGGCCTGAG ACTTTGTATT TCTACCAAGT TTCCAGGTGA
51   TGCTGAGGCT GCTGGCCCAG CGACCACATT TGATAATCAT AGCCCTCTGA
101  TAAATCCTAT CAAAATATCC TAATGGCAGA GCAAGGGAAT TCTGGTGATA
151  TCCTCCCCTA CCCATAACCT GACAGCTATT AGGATCTGCC TACTTGAGGC
201  TAAAAGCAAC CAAGAGAGGA ACAGCTACAG TGTACCACAG AGTCCCTCAA
251  CATCTTTGCC CACGCCACGG TGCCCCAGCT TCTTACCAAG TGTGCCTGAT
301  TCCTCTTGAC TACCTCCAAG GAAGTGGAGA AGGACAAGTT CTTGCGAAGC
351  CTTCGTCTTC TCTGATATGC TATTCTATGT CTATTTCTTT GGCCAAAAAG
401  ATGGGGCAAT GATATCAACT TTGCAGGGAG CTGGAGCATT TGCTAGTGAC
451  CTTTCTATGC CAGAACTTGC TAAGCATGCT AGCTAATAAT GATGTAGCAC
501  AGGGTGCGGT GGCTCACGCC TGTAATCTCA GCACTTTGGG CGGCCGAGGC
551  GGGCGGATCA CCTGAGGTCA GGAGTTCGAG ACCAGCCTGG CCAACATGAT
601  GAAACCCCAT CTCTACTAAA AATACAAAAA TTAGCCAGGC GTGGTGGTGG
651  GCACCTGCAA TCCCAGCTAC TCTGGAGGCT GAGACAGAAT CTCTTGAACC
701  CAGGAGGTGG AGATTGCAGT GAGCAGAGAT GGCACCACTG CATACCAGCC
751  TGGGCAACAA AGCAAGACTC TGTCTCAAAT AATAATAATA ATAATAACTA
801  ATGATGCAGC TTTCTCTCTC TGAGTATATA ATGCAGTTCT GATGATGTGA
851  GGAAGGGCCT CACTGTTGGT GTGGCAGAGA CTGACACCAT TGCTTGCAAT
901  GAAAACACTG CCCTTCGGTG CCTATGGGCT CTCCCTTTAT GGTTTCAGGG
951  AGGGCTTCTC AACCTTGGGA GAATTTTGGA CTGGATAGTT CTTTGTTGCA
1001 CAGGTGGGGG GCTGTCCTGC ACATCACAGG ATGTTTCATC CCTGGCCTCT
1051 ACCTACTAGA TGCCAGTAGA ACATACCCAC CCCACAGCTG CCTGTTGTGA
1101 CAATCAAAAG CATCTCCAGA TACTTTGCAG GGGGAAAATG ATTTCTCCAG
1151 GCCTGGCATA TACATAACAG TATTTAAGCA GCTGCCTAGA ATTAATTAAA
1201 CACAGAAGGA TGTCTCTCAT CCAGAATGCC CTGGACCACC TCTTTGATAG
1251 GCAATCAGAT CCCACCTCCT CCACCCTATT TTTGAAGGCC CTGTGCCAAC
1301 ACCACTTCTT CCATGAATAC TTCCTTGATT CCCCCATCCC TAGCTCTATA
1351 TAAATCTCCC ACTCAACACT CACACCTGTT AGTTTACATT CCTCTTGACA
1401 CTTGTCATTT AGCATCCTAA GTATGTAAAC ATGTCTCTCT TCACGATTCA
1451 CAAAGTGGCT TTGGAAGAAC TTTAGTACCT TCCCATCTTC TCTGCCATGG
1501 AAAGTGTACA CAACTGACAT TTTCTTTTTT TTTAAGACAG TATCTTGCTA
1551 TGATGGCCGG GCTGGAATGC TGTGGCTATT CACAGGCACA ATCATAGCTC
1601 ACTGCAGCCT TGAGCTCCCA GGCTCAAGTG ATCCTCCCGC CTCAGCCTCC
1651 TGAGTAGCTG AGATCACAGG CATGCACTAC CACACTCGGC TCACATTTGA
1701 CATCCTCTAA AGCATATATA AAATGTGGAG GAAAACTTTC ACAATTTGCA
1751 TCCCTTTGTA ATATGTAACA GAAATAAAAT TCTCTTTTAA AATCTATCAA
1801 CAATAGGCAA GGCACGGTGG CTCACGCCTG TCGTCTCAGC ACTTTGTGAG
1851 GCCCAGGCGG GCAGATCGTT TGAGCCTAGA AGTTCAAGAC CACCCTGGGC
1901 AACATAGCGA AACCCCCTTT CTACAAAAAA TACAAAAACT AGCTGGGTGT
1951 GGTGGTGCAC ACCTGTAGTC CCAGCTACTT GGAAGGCTGA AATGGGAAGA
```

2001 CTGCTTGAGC CCGGGAGGGG GAAGTTGCAG TAAGCCAGGA CCACACCACT
2051 GCACTCCAGC CTGGGCAACA GAGTGAGACT CTGTCTCAAA CAAACAAATA
2101 AATGAGGCGG GTGGATCACG AGGTCAGTAG ATCGAGACCA TCCTGGCTAA
2151 ACGGTGAAA CCCGTCTCTA CTAAAAAAAA AAAAAAAATA CAAAAAATTA
2201 CCAGGCATG GTGGCGGGCG CCTGTAGTCC CAGTTACTCG GGAGGCTGAG
2251 GCAGGAGAAT GGCGTGAAAC CGGGAGGCAG AGCTTGCAGT GAGCCGAGAT
2301 CGCACCACTG CCCTCCAGCC TGGGCGACAG AGCGAGACTC CGTCTCAATC
2351 AATCAATCAA TCAATAAAAT CTATTAACAA TATTTATTGT GCACTTAACA
2401 GGAACATGCC CTGTCCAAAA AAAACTTTAC AGGGCTTAAC TCATTTTATC
2451 CTTACCACAA TCCTATGAAG TAGGAACTTT TATAAAACGC ATTTTATAAA
2501 CAAGGCACAG AGAGGTTAAT TAACTTGCCC TCTGGTCACA CAGCTAGGAA
2551 GTGGGCAGAG TACAGATTTA CACAAGGCAT CCGTCTCCTG GCCCCACATA
2601 CCCAACTGCT GTAAACCCAT ACCGGCGGCC AAGCAGCCTC AATTTGTGCA
2651 TGCACCCACT TCCCAGCAAG ACAGCAGCTC CCAAGTTCCT CCTGTTTAGA
2701 ATTTTAGAAG CGGCGGGCCA CCAGGCTGCA GTCTCCCTTG GGTCAGGGGT
2751 CCTGGTTGCA CTCCGTGCTT TGCACAAAGC AGGCTCTCCA TTTTTGTTAA
2801 ATGCACGAAT AGTGCTAAGC TGGGAAGTTC TTCCTGAGGT CTAACCTCTA
2851 GCTGCTCCCC CACAGAAGAG TGCCTGCGGC CAGTGGCCAC CAGGGGTCGC
2901 CGCAGCACCC AGCGCTGGAG GGCGGAGCGG GCGGCAGACC CGGAGCAGCA
2951 TGTGGACTCT CGGGCGCCGC GCAGTAGCCG GCCTCCTGGC GTCACCCAGC
3001 CCGGCCCAGG CCCAGACCCT CACCCGGGTC CCGCGGCCGG CAGAGTTGGC
3051 CCCACTCTGC GGCCGCCGTG GCCTGCGCAC CGACATCGAT GCGACCTGCA
3101 CGCCCCGCCG CGCAGTAAGT ATCCGCGCCG GGAACAGCCG CGGGCCGCAC
3151 GCCGCGGGCC GCACGCCGCA CGCCTGCGCA GGGAGGCGCC GCGCACGCCG
3201 GGGTCGCTCC GGGTACGCGC GCTGGACTAG CTCACCCCGC TCCTTCTCAG
3251 GGTGGCCCGG CGGAAGCGGC CTTGCAACTC CCTTCTCTGG TTCTCCCGGT
3301 TGCATTTACA CTGGCTTCTG CTTTCCGAAG GAAAAGGGGA CATTTTGTCC
3351 TGCGGTGCGA CTGCGGGTCA AGGCACGGGC GAAGGCAGGG CAGGCTGGTG
3401 GAGGGGACCG GTTCCGAGGG GTGTGCGGCT GTCTCCATGC TTGTCACTTC
3451 TCTGCGATAA CTTGTTTCAG TAATATTAAT AGATGGTATC TGCTAGTATA
3501 TACATACACA TAATGTGTGT GTCTGTGTGT ATCTGTATAT AGCGTGTGTG
3551 TTGTGTGTGT GTGTTTGCGC GCACGGGCGC GCGCACACCT AATATTTTCA
3601 AGGCTGGATT TTTTTGAACG AAATGCTTTC CTGGAACGAG GTGAAACTTT
3651 CAGAGCTGCA GAATAGCTAG AGCAGCAGGG GCCCTGGCTT TTGGAAACTG
3701 ACCCGACCTT TATTCCAGAT TCTGCCCCAC TCCGCAGAGC TGTGTGACCT
3751 TGGGGGATTC CCCTAACCTC TCTGAGACGT GGCTTTGTTT TCTGTAGGGA
3801 GAAGATAAAG GTGACGCCCA TTTTGCGGAC CTGGTGTGAG GATTAAATGG
3851 GAATAACATA GATAAAGTCT TCAGAACTTC AAATTAGTTC CCCTTTCTTC
3901 CTTTGGGGGG TACAAAGAAA TATCTGACCC AGTTACGCCA CGGCTTGAAA
3951 GGAGGAAACC CAAAGAATGG CTGTGGGGAT GAGGAAGATT CCTCAAGGGG
4001 AGGACATGGT ATTTAATGAG GGTCTTGAAG ATGCCAAGGA AGTGGTAGAG
4051 GGTGTTTCAC GAGGAGGGAA CCGTCTGGGC AAAGGCCAGG AAGGCGGAAG
4101 GGGATCCCTT CAGAGTGGCT GGTACGCCGC ATGTATTAGG GGAGATGAAA
4151 GAGGCAGGCC ACGTCCAAGC CATATTTGTG TTGCTCTCCG GAGTTTGTAC
4201 TTTAGGCTTA AACTTCCCAC ACGTGTTATT TGGCCCACAT TGTGTTTGAA

4251 GAAACTTTGG GATTGGTTGC CAGTGCTTAA AAGTTAGGAC TTAGAAAATG
4301 GATTTCCTGG CAGGACGCGG TGGCTCATGC CCATAATCTC AGCACTTTGG
4351 GAGGCCTAGG AAGGTGGATC ACCTGAGGTC CGGAGTTCAA GACTAACCTG
4401 GCCAACATGG TGAAACCCAG TATCTACTAA AAAATACAAA AAAAAAAAAA
4451 AAAAAGAAGA AGAAGAAGAA GAAGAAGAAG AAAATAAAGA AAAGTTAGCC
4501 GGGCGTGGTG TCGCGCGCCT GTAATCCCAG CTACTCCAGA GGCTGCGGCA
4551 GGAGAATCGC TTGAGCCCGG GAGGCAGAGG TTGCATTAAG CCAAGATCGC
4601 CCAATGCACT CCGGCCTGGG CGACAGAGCA AGACTCCGTC TCAAAAAATA
4651 ATAATAATAA ATAAAAATAA AAAATAAAAT GGATTTCCCA GCATCTCTGG
4701 AAAAATAGGC AAGTGTGGCC ATGATGGTCC TTAGATCTCC TCTAGGAAAG
4751 CAGACATTTA TTACTTGGCT TCTGTGCACT ATCTGAGCTG CCACGTATTG
4801 GGCTTCCACC CCTGCCTGTG TGGACAGCAT GGGTTGTCAG CAGAGTTGTG
4851 TTTTGTTTTG TTTTTTTGAG ACAGAGTTTC CCTCTTGTTG CCCAGGCTGG
4901 AGTGCAGTGG CTCAGTCTCA GCTCACTGCA ACCTCTGCCT CCTGGGTTCA
4951 AGTGATTCTC CTGCCTCAGC CTCCCGAGTA GCTGGGATTA TCGGCTAATT
5001 TTGTATTTTT AGTAGAGACA GATTTCTCCA TGTTGGTCAG GCTGGTCTCG
5051 AACTCCCAAC CTCAGGTGAT CCGCCCACCT CGCCCTCCCA AAGTGCTGGA
5101 ATTACAGGCG TGAGCCACCG CGTCTGGCCA TCAGCAGAGT TTTTAATTTA
5151 GGAGAATGAC AAGAGGTGGT ACAGTTTTTT AGATGGTACC TGGTGGCTGT
5201 TAAGGGCTAT TGACTGACAA ACACACCCAA CTTGGCGCTG CCGCCCAGGA
5251 GGTGGACACT GGGTTTCTGG ATAGATGGTT AGCAACCTCT GTCACCAGCT
5301 GGGCCTCTTT TTTTCTATAC TGAATTAATC ACATTTGTTT AACCTGTCTG
5351 TTCCATAGTT CCCTTGCACA TCTTGGGTAT TTGAGGAGTT GGGTGGGTGG
5401 CAGTGGCAAC TGGGGCCACC ATCCTGTTTA ATTATTTTAA AGCCCTGACT
5451 GTCCTGGATT GACCCTAAGC TCCCCCTGGT CTCCAAAATT CATCAGAAAC
5501 TGAGTTCACT TGAAGGCCTC TTCCCCACCC TTTTCTCCAC CCCTTGCATC
5551 TACTTCTAAA GCAGCTGTTC AACAGAAACA GAATGGGAGC CACACACATA
5601 ATTCTACATT TTCTAGTTAA AAAGAAAAAA AAATCATTTT CAACAATATA
5651 TTTATTCAAC CTAGTACATA CAAAATATTA TCATTCCAAC ATGTAATCAG
5701 TATTTTAAAA ATCAGTAATG AGACCAGGCA CGGTGGCTCC CGACTGTAAT
5751 CCCAGGACTT TGGGAGGCCG AGGCGAGTGG ATCATCTGAG ATCAGGAGTT
5801 CAAGACCAGC CTGGCCAACA TGGTGAAACC CCATCTCTAC TAAACACTAG
5851 CTCAGCATGG TGGTGGGTGC CTGTAGTCCC AGCTACTCGG GAGGCTGAGG
5901 CATGAGAATC ACTTGAGCCC AGGAGGCAGA GGTTGCAGTG AGCCAAGATT
5951 TTGGGGGATT CTGTGACATA CAAAAAAAAT CAGTAATAAG ATATCTTGCA
6001 TACTCTTTTC GTACTCATAT ACTTCCAGCA TATCTCAATT CACAATTTCT
6051 AAGTAAATGC TCTATCTGTA TTTACTTTTA TAAAATTCAC AATTAAAAAT
6101 GAAGGTTCAC ATAGTCAAGT TGTTCCAAAC ACACTTAAAT GTCTCCTAGG
6151 CTGGGTGTGG TTGCTCACAC CTGTAATCCC AGCACTTTGG GAGGCTGAGA
6201 TGGGCGGATC ACCTGAGGTC AGGAGTTTGA GACCAGCCTG GCCAACATGG
6251 TGAAACCCCG TCTCTACTAA AAATACAAAA ATTAGCTGGA TGTGGTGGCA
6301 CTCACATGTA ATCCCAGCTA CTCAGGAGGC TGAGGCAGGA TAATTGCTTG
6351 AACCCGGGAG GTGGTGGAGG TTGCAGTGAG CCGAGATCGC ACCACTGCCT
6401 TCCAACCTGG GCGACAGAGC GAGACTCCGT CTCAAAAAAA AAAAAAAGGC
6451 TCCTAATAAC TTTATTACTT TATTATCACC TCAAATAATT AAAATTAAAT

6501 GAAGTTGAAA ATCCAGGTCC TCAGTCCCAT TAGCCACATT TCTAGTGCTC
6551 AGTAGCCACG GGGGCTGGTG ACCACCACAT GGGACAGCAT ATTTAGTACC
6601 TGATCATTGG TTCTCAGATC TGGCTACTCA GCAGAACCAA GAATCCACAG
6651 AAACGGCTTT TAAAAGCACA GCCCCACAGC CCCCAGCCCC AGCCTTACTA
6701 CCTGGAGGCT GGGAAGGACT CTGATTCCAC GAGGCAGCCT ATGTTTTTTG
6751 ATGGAGGGAT GTGACAGGGG CTGCATCTTT AACGTTTCCT CTTAAATACT
6801 GGAGACAGCT TCGAGGAGGA GATAACTGGA TGTGTCTTAG TCCATTTGAT
6851 GGAGGGATGT GACGGGGCTG CGTCTTTAAC GTTTCCTCTT AAATACCGGA
6901 GACAGCTTCG AGAAGGAGAT AACTGGATGT TTCTTAGTCC ACTTTCTGTT
6951 GCTTGTGACA GAATACCTGA AACTGGGCAA TTTATATGGT AAAAAATTTT
7001 CTTCTTACTG CTCTGGAGGC TGAGAAGTCC AAAGTCAAGT CCCTTCTTGC
7051 TGGTGGGGAC TTTGCAGAGT ATTGAGGCGG CACCGGGCGT CATATGGTAA
7101 GGGGCTGAGT GTGCTACCTC AGGTGTCTTT TTCTTTTCTT ATAAAGCCTA
7151 ACTAGTTTCA CTCCCATGAT AACCCATTAA TCTATGAATG GATTAATCCA
7201 TTATTGAGGG AAGAACCTTC ATGACCCAGT CACCGCTTAA AGGCCCCACC
7251 TCTCAATACT GCCACATCGG GAATTAAGTT TCAACATGAG TTTCGGAGGT
7301 GACAAACATT CAAACCATAG CATGCTGTCT CTTAAATGAC TCAATAAGCT
7351 CCTGTGGCAT CCACTTCTGC ATGCCTTGGG CAGCTTTTAG ACATCTGTCC
7401 ATTTTCCTAG AGGGACAAGA CCACCACCTG TGATCCTATG ACCTTTTGGC
7451 TTTAGGCCTA ACAAGCAGGT TATACCCTCA CTCACTTTCA AATCATTTTT
7501 ATTGTCTTGC AGACAATTTA CACAAGTTTA CACATAGAAA AGGATATGTA
7551 AATATTTATA CGCTGCCGGG CGCGGTGGCT CACGCCTGTA ATCCCAGCAC
7601 TTTGGGAGGC CGAGGCAGGT GGATCACGAG TTCAGGAGAT GGAGACCATC
7651 CTGGCTAATA CGATGAAACC CCATCTCTAC TAAAAATACA AAAAATTAGC
7701 CGGGCGTGGT GACGGGTGCC TGTAGTCCCC ACTACTCGGG ACGCTGAGGC
7751 AGGAGAATGG CGTGAACCCG GGAGGCAGAG CTTGCAGTGA TCCGAGATCG
7801 TGCCACTGCA CTCCAGCCTG GGTGACAGAG CGAGACTGCA TCTCAAAGAA
7851 AAAAATAAAT AAATAAATAA ATATTTATAC TGCTTATAAA CTAATAATAA
7901 ATGCTATGGT CTGCATGTTT GTGTCACCCC ACCATTCATA TGTTAAAACC
7951 TAATCACCAA AGTGATATTA GGAGGTGGGG CCCTTGGGAG GTGATGAGGT
8001 ATGAGGGTGG AGCCCATATG ATTGGGATTA GTGCCCTTCT AAAATAGCCC
8051 AACGGAGCCC AGTGACAAGG CATCATCTAT GAACCAGGAA ACTGGCCCTC
8101 ACCAGACACC AAAGCTGTTG GTGCATTGAT CTTGGATTTC CCACCCTCCA
8151 GGACTCTAAG AAACACATTT CTATTGTTTA TAAGCCACCC AGTGGCTGGT
8201 ATTTTGTTAT AACATCCCAG ACTAAGACAA ATAACAAATA CTTGTATCCC
8251 TGACACCAGG TTAAGAGATA GAATTTGTTT GTTCCTCTGG AGGCCCTTGT
8301 CTTCACCCCA TCACTGCCCT GTCCTCCCTG GAGGAATCTG CCAGCCCGAA
8351 TTC

SEQUENCE ID 2 (SEQ ID NO. 2)

1 TTTACAGGGC ATAACTCATT TTATCCTTAC CACAATCCTA TGAAGTAGGA
51 ACTTTTATAA AACGCATTTT ATATNCAAGG GCACAGAGAG GNTAATTAAC
101 TTGCCCTCTG GTCACACAGC TAGGAAGTGG GCAGAGTACA GATTTACACT
151 AGGCATCCGT CTCCTGNCCC CACATANCCA GCTGCTGTAA ACCCATACCG
201 GCGGCCAAGC AGCCTCAATT TGTGCATGCA CCCACTTCCC AGCAAGACAG
251 CAGCTCCCAA GTTCCTCCTG TTTAGAATTT TAGAAGCGGC GGGCCACCAG
301 GCTGCAGTCT CCCTTGGGTC AGGGGTCCTG GTTGCACTCC GTGCTTTGCA
351 CAAAGCAGGC TCTCCATTTT TGTTAAATGC ACGAATAGTG CTAAGCTGGG
401 AAGTTCTTCC TGAGGTCTAA CCTCTAGCTG CTCCCCCACA GAAGAGTGCC
451 TGCGGCCAGT GGCCACCAGG GGTCGCCGCA GCACCCAGCG CTGGAGGGCG
501 GAGCGGGCGG CAGACCCGGA GCAGCATGTG GACTCTCGGG CGCCGCGCAG
551 TAGCCGGCCT CCTGGCGTCA CCCAGCCCGG CCCAGGCCCA GACCCTCACC
601 CGGGTCCCGC GGCCGGCAGA GTTGGCCCCA CTCTGCGGCC GCCGTGGCCT
651 GCGCACCGAC ATCGATGCGA CCTGCACGCC CCGCCGCGCA AGTTCGAACC
701 AACGTGGCCT CAACCAGATT TGGAATGTCA AAAAGCAGAG TGTCTATTTG
751 ATGAATTTGA GGAAATCTGG AACTTTGGGC CACCCAGGCT CTCTAGATGA
801 GACCACCTAT GAAAGACTAG CAGAGGAAAC GCTGGACTCT TTAGCAGAGT
851 TTTTTGAAGA CCTTGCAGAC AAGCCATACA CGTTTGAGGA CTATGATGTC
901 TCCTTTGGGA GTGGTGTCTT AACTGTCAAA CTGGGTGGAG ATCTAGGAAC
951 CTATGTGATC AACAACAGAC GCCAAACAAG CAAATCTGGC TATCTTCTCC
1001 ATCCAGTGGA CCTAAGCGTT ATGACTGGAC TGGGAAAAAC TGGGTGTTCT
1051 CCCACGACGG CGTGTCCCTC CATGAGCTGC TGGCCGCAGA GCTCACTAAA
1101 GCCTTAAAAA CCAAACTGGA CTTGTCTTGG TTGGCCTATT CCGGAAAAGA
1151 TGCTTGATGC CCAGCCCCGT TTTAAGGACA TTAAAAGCTA TCAGGCCAAG
1201 ACCCCAGCTT CATTATGCAG CTGAGGTGTG TTTTTTGTTG TTGTTGTTGT
1251 TTATTTTTTT TATTCCTGCT TTTGAGGACA CTTGGGCTAT GTGTCACAGC
1301 TCTGTACAAA CAATGTGTTG CCTCCTACCT TGCCCCCAAG TTCTGATTTT
1351 TAATTTCTAT GGAAGATTTT TTGGATTGTC GGATTTCCTC CCTCACATGA
1401 TACCCCTTAT CTTTTATAAT GTCTTATGCC TATACCTGAA TATAACAACC
1451 TTTAAAAAAG CAAAATAATA AGAAGGAAAA ATTCCAGGAG GGA

SEQUENCE ID NO. 3 (SEQ ID NO. 3):

1 ATGTGGACTC TCGGGCGCCG CGCAGTAGCC GGCCTCCTGG CGTCACCCAG
51 CCCGGCCCAG GCCCAGACCC TCACCCGGGT CCCGCGGCCG GCAGAGTTGG
151 CCCCACTCTG CGGCCGCCGT GGCCTGCGCA CCGACATCGA TGCGACCTGC
201 ACGCCCCGCC GNGCAAGTTC GAACCAACGT GGCCTCAACC AGATTTGGAA
251 TGTCAAAAAG CAGAGTGTCT ATTTGATGAA TTTGAGGAAA TCTGGAACTT
301 TGGGCCACCC AGGCTCTCTA GATGAGACCA CCTATGAAAG ACTAGCAGAG
351 GAAACGCTGG ACTCTTTAGC AGAGTTTTTT GAAGACCTTG CAGACAAGCC
401 ATACACGTTT GAGGACTATG ATGTCTCCTT TGGGAGTGGT GTCTTAACTG
451 TCAAACTGGG TGGAGATCTA GGAACCTATG TGATCAACAA GCAGACGCCA
501 AACAAGCAAA TCTGGCTATC TTCTCCATCC AGGTTAACGT GGCTCCTGTG
551 GCTGTTCCAT CCCTGAGGAA AAGTGAGGAC CATGCTCTCC AAACAGGCCA
601 TGTGCTGGAC TACCTCTGTT TCTGTCTCCT GGGATTCCAA TCAGCAAGTG
651 AGCAACGAAG CAACCCAGAC AGTGTGGTTC ATAGGATGGC TGG

SEQUENCE ID NO 4 (SEQ ID NO. 4):

1 MWTLGRRAVA GLLASPSPAQ AQTLTRVPRP AELAPLCGRR GLRTDIDATC
51 TPRRASSNQR GLNQIWNVKK QSVYLMNLRK SGTLGHPGSL DETTYERLAE
101 ETLDSLAEFF EDLADKPYTF EDYDVSFGSG VLTVKLGGDL GTYVINKQTP
151 NKQIWLSSPS SGPKRYDWTG KNWVFSHDGV SLHELLAAEL TKALKTKLDL
201 SWLAYSGKDA

SEQUENCE ID NO 5 (SEQ ID NO. 5)

1 MWTLGRRAVA GLLASPSPAQ AQTLTRVPRP AELAPLCGRR GLRTDIDATC
51 TPRRASSNQR GLNQIWNVKK QSVYLMNLRK SGTLGHPGSL DETTYERLAE
101 ETLDSLAEFF EDLADKPYTF EDYDVSFGSG VLTVKLGGDL GTYVINKQTP
151 NKQIWLSSPS RLTWLLWLFH P

SEQUENCE ID NO 6 (SEQ ID NO. 6).

1 CAAGCCTGGG CGACAGAGCG AGCTCCGTCN CAACCAATNA ACCAATCAAT AAAATCTANN
61 AACAATATTT ATTGTGCACT TAACAGGAAC ATGCCCTGTC CAAAAAAAAC TTTACAGGGC
121 TTAACTCATT TTATCCTTAC CACAATCCTA TGAAGTAGGA ACTTTTATAA AACGCATTTT
181 ATAAACAAGG CACAGAGAGG TTAATTAACT TGCCCTCTGG TCACACAGCT AGGAAGTGGG
241 CAGAGTACAG ATTTACACAA GGCATCCGTC TCCTGGCCCC ACATACCCAA CTGCTGTAAA
301 CCCATACCGG CGGCCAAGCA GCCTCAATTT GTGCATGCAC CCACTTCCCA GCAAGACAGC
361 AGCTCCCAAG TTCCTCCTGT TTAGAATTTT AGAAGCGGCG GGCCACCAGG CTGCAGTCTC
421 CCTTGGGTCA GGGGTCCTGG TTGCACTCCG TGCTTTGCAC AAAGCAGGCT CTCCATTTTT
481 GTTAAATGCA CGAATAGTGC TAAGCTGGGA AGTTCTTCCT GAGGTCTAAC CTCTAGCTGC
541 TCCCCCACAG AAGAGTGCCT GCGGCCAGTG GCCACCAGGG GTCGCCGCAG CACCCAGCGC
601 TGGAGGGCGG AGCGGGCGGC AGACCCGGAG CAGCATGTGG ACTCTCGGGC GCCGCGCAGT
661 AGCCGGCCTC CTGGCGTCAC CCAGCCCGGC CCAGGCCCAG ACCCTCACCC GGGTCCCGCG
721 GCCGGCAGAG TTGGCCCCAC TCTGCGGCCG CCGTGGCCTG CGCACCGACA TCGATGCGAC
781 CTGCACGCCC CGCCGNGCAG TAAGTATCCG CGCCGGGAAC AGCCGCGGGC CGCACGCCGY
841 GGGCCGCACG CCGCACGCCT GCGCAGGGAG GCGCCGCGCA CGCCGGGGTC GCTCCGGGTA
901 CGCGCGCTGG ACTAGCTCAC CCCGCTCCTT CTCAGGGTGG CCCGGCGGAA GCGGCCTTGC
961 AACTCCCTTC TCTGGTTCTC CCGGTTGCAT TTACACTGGC TTCTGCTTTC CGAAGGAAAA
1021 GGGGACATTT TGTCCTGCGG TGCGACTGCG GGTCAAGGCA CGGGCGAAGG CAGGGCAGGC
1081 TGGTGGAGGG GACCGGTTCC GAGGGGTGTG CGGCTGTCTC CATGCTTGTC ACTTCTCTGC
1141 GATAACTTGT TTCAGTAATA TTAATAGATG GTATCTGCTA GTATATACAT ACACATAATG
1201 TGTGTGTCTG TGTGTATCTG TATATAGCGT GTGTGTTGTG TGTGTGTGTT TGCGCGCACG
1261 GGCGCGCGCA CACCTAATAT TTTCAAGGCT GGATTTTTTT GAACGAAATG CTTTCCTGGA
1321 ACGAGGTGAA ACTTTCAGAG CTGCAGAATA GCTAGAGCAG CAGGGGCCCT GGCTTTTGGA
1381 AACTGACCCG ACCTTTATTC CAGATTCTGC CCCACTCCGC AGAGCTGTGT GACCTTGGGG
1441 GATTCCCCTA ACCTCTCTGA GACGTGGCTT TGTTTTCTGT AGGGAGAAGA TAAAGGTGAC
1501 GCCCATTTTG CGGACCTGGT GTGAGGATTA AATGGGAATA ACATAGATAA AGTCTTCAGA
1561 ACTTCAAATT AGTTCCCCTT TCTTCCTTTG GGGGGTACAA AGAAATATCT GACCCAGTTA
1621 CGCCACGGCT TGAAAGGAGG AAACCCAAAG AATGGCTGTG GGGATGAGGA AGATTCCTCA
1681 AGGGGAGGAC ATGGTATTTA ATGAGGGTCT TGAAGATGCC AAGGAAGTGG TAGAGGGTGT
1741 TTCACGAGGA GGGATCCGTC TGGGCAAAGG CCAGGAAGGC GGAAGGGGAT CCCTTCCGAG
1801 TGGCTGGTAC GCCGCCTGTA NTATGGGAGA GGATCCCTTC AGAGTGGCTG GTACGCCGCA
1861 TGTATTAGGG GAGATGAAAG AGGCAGGCCA CGTCCAAGCC ATATTTGTGT TGCTCTCCGG
1921 AGTTTGTACT TTAGGCTTAA ACTTCCCACA CGTGTTATTT GGCCCACATT GTGTTTGAAG
1981 AAACTTTGGG ATTGGTTGCC AGTGCTTAAA AGTTAGGACT TAGAAAATGG ATTTCCTGGC
2041 AGGACGCGGT GGCTCATGCC CATAATCTCA GCACTTTGGG AGGCCTAGGA AGGTGGATCA
2101 CCTGAGGTCC GGAGTTCAAG ACTAACCTGG CCAACATGGT GAAACCCAGT ATCTACTAAA
2161 AAATACAAAA AAAAAAAAAA AAAAGAAGAA GAAGAAGAAG AAGAAGAAGA AAATAAAGAA
2221 AAGTTAGCCG GGCGTGGTGT CGCGCGCCTG TAATCCCAGC TACTCCAGAG GCTGCGGCAG
2281 GAGAATCGCT TGAGCCCGGG AGGCAGAGGT TGCATTAAGC CAAGATCGCC CAATGCACTC
2341 CGGCCTGGGC GACAGAGCAA GCTCCGTCTC AAAAAATAAT AATAATAAAT AAAAATAAAA
2401 AATAAAATGG ATTTCCCAGC ATCTCTGGAA AAATAGGCAA GTGTGGCCAT GATGGTCCTT
2461 AGATC

SEQUENCE ID NO 7 (SEQ ID NO. 7)

1 AATTTACTCC GAAACTAGCT TGGGTGAGGG GTACAAAGCA TCCTGCCTTT CTTTAAAAGT
61 GCTGCTTCCC CTTGGAAGTA GAAAGTGGAC ACTTTTATAA GGTAAGGGGG GAAGTGTGCA
121 AGGGCAAGTG GGGGGGTCCC TCTGCTAGTT CCGTGCATAC TCTACAGGAC AGTTGACTTG
181 GCACCTTCCT GGTTAGTAAT AAGCTGTAGC AGTGGCCAAG TGGGCATGCT TTCAGTATGC
241 CCTCCCAGTG AATGAAAGTC CTGAGGCAAC CCCCAAGGGT GGAAGTGCCA GGCCACCACC
301 CACTGGAGGT GAAAGTTCCG TGATGGGTTT GCTTTGGTCT GCGAATCTAC TGTCATGTGG
361 AGAGATCTGT GCTCTGGAAG AGCATACAGT TAGAAAAGCT TGCCCTGAAG GGAATGTATG
421 GTGAAGGGGA GGTGAAAGGT TATATTTGCA TTTCTGAAGG GCTAAGTAGG AAACCGGGAA
481 CCAGGGGAGA GGAGAAGAGA AGAGAGGATA ATTTTTTTTA AGAAAAGCAA CATATTCCCT
541 TTTTCTTAGA AAAAATGGAG CACTCGGTTA CAGGCACTCG AATGTAGAAG TAGCAATATA
601 TAAATTATGC ATTAATGGGT TATAATTCAC TGAAAAATAG TAACGTACTT CTTAACTTTG
661 GCTTTCAGAGTTTCGAACCAA CGTGGCCTCA ACCAGATTTG GAATGTCAAA AAGCAGAGTG
721 TCTATTTGAT GAATTTGAGG AAATCTGGAA CTTTGGGCCA CCCAGGGTAA GATAAAGCAN
781 CTTNCACGTG ATAGGTATCT TCCTCTNTCC TTCCCTGCCT CTCCCATTAG AACCTGGTTT
841 TCTTTCTGAG CAGCAACAAT NTTAGGCATC TTTCCATGTG ACTGAGTATC CACCACATTA
901 TTTTTAATGA AATAGTATTA GATTGCATGG ATGTGACATA ATCCATTTAA CNGATCNCCT
961 ACTGTTGGAC ATTCAGGTTG TTTTCAGAGT TTNATATTAT TTTATTTAAT ACCCTAATAG
1021 TTAGAGCAGG CCATGCTTNT NTTACAAATA GGACCCAAAT ATTTAATAGC TCAAACCAAT
1081 AACGGTNTGT GTCCTCCTCT CTGGGCAGTA CAGGGTTGGC ATACCTCTGA AGTGATTAGG
1141 GNCTACACTC ATTCNAGCTT CCAGTTGGCC TTATCTGTCA GTGCCTACT

SEQUENCE ID NO 8 (SEQ ID NO. 8):

1 AAAATGGAAG CATTTGGTAA TCATGTTTGG GTTTTGTGCT TCCTCTGCAG CTCTCTAGAT
61 GAGACCACCT ATGAAAGACT AGCAGAGGAA ACGCTGGACT CTTTAGCAGA GTTTTTTGAA
121 GACCTTGCAG ACAAGCCATA CACGTTTGAG GACTATGATG TCTCCTTTGG GGTACCTCTT
181 GACTTCTTTT ATTTTTCTGT TTCCCCCTCT AAGAATTTTA GTTCACT

SEQUENCE ID NO 9 (SEQ ID NO. 9):

1 AAGCAATGAT GACAAAGTGC TAACTTTTTC TTGTTTTAAT TTCTTTATGC TTTTTTTCCA
61 CCTAATCCCC TAGAGTGGTG TCTTAACTGT CAAACTGGGT GGAGATCTAG GAACCTATGT
121 GATCAACAAG CAGACGCCAA ACAAGCAAAT CTGGCTATCT TCTCCATCCA GGTATGTAGG
181 TATGTTCAGA AGTCAACATA TGTAATTCTT AAAGACTTCC GAAATGTGAC ATTGTGGACC
241 A

SEQUENCE ID NO 10 (SEQ ID NO. 10)

1 TCATCTGAAG GGCTGTGCTG TGGAATTACT ATGCATTTGT TTTGTCTTCC AGTGGACCTA
61 AGCGTTATGA CTGGACTGGG AAAAACTGGG TGTTCTCCCA CGACGGCGTG TCCCTCCATG
121 AGCTGCTGGC CGCAGAGCTC ACTAAAGCCT TAAAAACCAA ACTGGACTTG TCTTGGTTGG
181 CCTATTCCGG AAAAGATGCT TGATGCCCAG CCCCGTTTTA AGGACATTAA AAGCTATCAG
241 GCCAAGACCC CAGCTTCATT ATGCAGCTGA GGTGTGTTTT TTGTTGTTGT TGTTGTTTAT
301 TTTTTTTATT CCTGCTTTTG AGGACACTTG GGCTATGTGT CACAGCTCTG TACAAACAAT
361 GTGTTGCCTC CTACCTTGCC CCCAAGTTCT GATTTTTAAT TTCTATGGAA GATTTTTTGG
421 ATTGTCGGAT TTCCTCCCTC ACATGATACC CCTTATCTTT TATAATGTCT TATGCCTATA
481 CCTGAATATA ACAACCTTTA AAAAAGCAAA ATAATAAGAA GGAAAAATTC CAGGAGGG

SEQUENCE ID NO 11 (SEQ ID NO. 11):

1 CCTAGGAGGT GTAGCCTGGG AACCATAGGC AAGAATAATT AACTCAGCTC CTCGGTTAGT
61 GCCTCCTCAG TTCGAGATGG AATTTATTTG CAGGCATGGC TCCTTAATAT GCCAAACCCA
121 TGCTCAAGAC ATACTCCTTC TCCTGGAAGG TTAACGTGGC TCCTGTGGCT GTTCCATCCC
181 TGAGGAAAAG TGAGGACCAT GCTCTCCAAA CAGGCCATGT GCTGGACTAC CTCTGTTTCT
241 GTCTCCTGGG ATTCCAATCA GCAAGTGAGC AACGAAGCAA CCCAGACAGT GTGGTTCATA
301 GGATGGCTGG GTAAGTGGCT GTTTGTTTTT TCCTTACTGT GGATATGTAT CAGTGAAGGA
361 ATCTGTAGAA CATTCTTGAT GGGAACATTT AGTCATATCA AGTCAATAAA TTAATGTTTA
421 GGCTGGGAC

SEQUENCE ID NO 12 (SEQ ID NO. 12):

1 TTTACAGGGC ATAACTCATT TTATCCTTAC CACAATCCTA TGAAGTAGGA ACTTTTATAA
61 AACGCATTTT ATATNCAAGG GCACAGAGAG GNTAATTAAC TTGCCCTCTG GTCACACAGC
121 TAGGAAGTGG GCAGAGTACA GATTTACACT AGGCATCCGT CTCCTGNCCC CACATANCCA
181 GCTGCTGTAA ACCCATACCG GCGGCCAAGC AGCCTCAATT TGTGCATGCA CCCACTTCCC
241 AGCAAGACAG CAGCTCCCAA GTTCCTCCTG TTTAGAATTT TAGAAGCGGC GGGCCACCAG
301 GCTGCAGTCT CCCTTGGGTC AGGGGTCCTG GTTGCACTCC GTGCTTTGCA CAAAGCAGGC
361 TCTCCATTTT TGTTAAATGC ACGAATAGTG CTAAGCTGGG AAGTTCTTCC TGAGGTCTAA
421 CCTCTAGCTG CTCCCCCACA GAAGAGTGCC TGCGGCCAGT GGCCACCAGG GGTCGCCGCA
481 GCACCCAGCG CTGGAGGGCG GAGCGGGCGG CAGACCCGGA GCAGCATGTG GACTCTCGGG
541 CGCCGCGCAG TAGCCGGCCT CCTGGCGTCA CCCAGCCCGG CCCAGGCCCA GACCCTCACC
601 CGGGTCCCGC GGCCGGCAGA GTTGGCCCCA CTCTGCGGCC GCCGTGGCCT GCGCACCGAC
661 ATCGATGCGA CCTGCACGCC CCGCCGCGCA AGTTCGAACC AACGTGGCCT CAACCAGATT
721 TGGAATGTCA AAAAGCAGAG TGTCTATTTG ATGAATTTGA GGAAATCTGG AACTTTGGGC
781 CACCCAGGCT CTCTAGATGA GACCACCTAT GAAAGACTAG CAGAGGAAAC GCTGGACTCT
841 TTAGCAGAGT TTTTTGAAGA CCTTGCAGAC AAGCCATACA CGTTTGAGGA CTATGATGTC
901 TCCTTTGGGA GTGGTGTCTT AACTGTCAAA CTGGGTGGAG ATCTAGGAAC CTATGTGATC
961 AACAAGCAGA CGCCAAACAA GCAAATCTGG CTATCTTCTC CATCCAGTGG ACCTAAGCGT
1021 TATGACTGGA CTGGGAAAAA CTGGGTGTTC TCCCACGACG GCGTGTCCCT CCATGAGCTG
1081 CTGGCCGCAG AGCTCACTAA AGCCTTAAAA ACCAAACTGG ACTTGTCTTG GTTGGCCTAT
1141 TCCGGAAAAG ATGCTTGATG CCCAGCCCCG TTTTAAGGAC ATTAAAAGCT ATCAGGCCAA
1201 GACCCCAGCT TCATTATGCA GCTGAGGTGT GTTTTTTGTT GTTGTTGTTG TTTATTTTTT
1261 TTATTCCTGC TTTTGAGGAC ACTTGGGCTA TGTGTCACAG CTCTGTACAA ACAATGTGTT
1321 GCCTCCTACC TTGCCCCCAA GTTCTGATTT TTAATTTCTA TGGAAGATTT TTTGGATTGT
1381 CGGATTTCCT CCCTCACATG ATACCCCTTA TCTTTTATAA TGTCTTATGC CTATACCTGA
1441 ATATAACAAC CTTTAAAAAA GCAAAATAAT AAGAAGGAAA AATTCCAGGA GGGAAAAAAA
1501 AAAAA

SEQUENCE ID NO 13 (SEQ ID NO. 13).

1 CCAAAGTTCC AGATTTCCTC A

SEQUENCE ID NO 14 (SEQ ID NO. 14):

1 TCCCGCGGCC GGCAGAGTT

SEQUENCE ID NO 15 (SEQ ID NO. 15).

1 AGCACCCAGC GCTGGAGG

SEQUENCE ID NO 16 (SEQ ID NO. 16):

1 CCGCGGCTGT TCCCGG

SEQUENCE ID NO 17 (SEQ ID NO. 17):

1 AGTAACGTAC TTCTTAACTT TGGC

SEQUENCE ID NO 18 (SEQ ID NO 18):

1 AGAGGAAGAT ACCTATCACG TG

SEQUENCE ID NO 19 (SEQ ID NO 19):

1 AAAATGGAAG CATTTGGTAA TCA

SEQUENCE ID NO 20 (SEQ ID NO. 20):

1 AGTGAACTAA AATTCTTAGA GGG

SEQUENCE ID NO 21 (SEQ ID NO 21):

1 AAGCAATGAT GACAAAGTGC TAAC

SEQUENCE ID NO 22 (SEQ ID NO. 22):

1 TGGTCCACAA TGTCACATTT CGG

SEQUENCE ID NO 23 (SEQ ID NO. 23):

1 CTGAAGGGCT GTGCTGTGGA

SEQUENCE ID NO 24 (SEQ ID NO. 24):

1 TGTCCTTACA AACGGGGCT

SEQUENCE ID NO 25 (SEQ ID NO. 25):

1 CCCATGCTCA AGACATACTC C

SEQUENCE ID NO 26 (SEQ ID NO. 26):

1 ACAGTAAGGA AAAAACAAAC AGCC

SEQUENCE ID NO 27 (SEQ ID NO. 27):

1 GGGCTGGCAG ATTCCTCCAG

SEQUENCE ID NO 28 (SEQ ID NO. 28):

1 GTAAGTATCC GCGCCGGGAA C

SEQUENCE ID NO 29 (SEQ ID NO. 29):

1 GGGATTGGTT GCCAGTGCTT AAAAGTTAG

SEQUENCE ID NO 30 (SEQ ID NO. 30):

1 GATCTAAGGA CCATCATGGC CACACTTGCC

SEQUENCE ID NO 31 (SEQ ID NO. 31)

1 GGAGGGATCC GTCTGGGCAA AGG

SEQUENCE ID NO 32 (SEQ ID NO. 32):

1 CAATCCAGGA CAGTCAGGGC TTT

SEQUENCE ID NO 33 (SEQ ID NO. 33).

1 TCCCGCGGCC GGCAGAGTT

What is claimed is:

1. A method of screening individuals for a mutation that leads to Friedreich's ataxia, comprising the steps of:

digesting DNA from an individual to be tested with a restriction endonuclease; and

measuring the length of a restriction fragment length polymorphism (RFLP) by hybridization to probes that recognize a region encompassing a GAA repeat in the first intron of an X25 gene and performing Southern Blot analysis, wherein an RFLP having said GAA expansion of more than about 120 is an indication of said mutation that leads to Friedreich's ataxia.

2. The method of claim 1, wherein the restriction endonuclease is EcoRI.

3. The method of claim 1, wherein the probe used for performing said Southern Blot is SEQ ID NO 2.

4. The method of claim 1, wherein the probe used for performing said Southern Blot is an amplification product obtained by performing PCR on said DNA with SEQ ID NO 16 and SEQ ID NO 17.

5. A method of screening individuals for a mutation that leads to Friedreich's ataxia, comprising the steps of measuring expression of an X25 gene by determining an amount of mRNA expressed from said X25 gene and from known controls, and comparing the amount of mRNA from said X25 gene to the amount of mRNA from the known controls.

6. The method of claim 5, wherein the mRNA is determined by the steps of:

extracting mRNA from individuals to be tested;

preparing cDNA from mRNA;

amplifying said cDNA to produce amplification products; and

comparing relative amounts of X25 and control amplification products present, wherein a reduced amount of mRNA from the X25 gene indicates individuals having said mutation that leads to Friedreich's ataxia.

7. The method of claim 6, wherein the comparing step includes electrophoresis of said amplification products; transfering said amplification products to a solid support; hybridizing said amplification products to a probe; and quantifying of X25 amplification products versus control gene amplification products.

8. The method of claim 6, wherein said probe is SEQ ID NO 14.

9. The method of claim 5, wherein said control gene is serine hydroxymethyltransferase (SHMT).

10. A method of screening individuals for a mutation that leads to Friedreich's ataxia, comprising the step of detecting a variation in a size of a $(GAA)_n$ repeat in a first intron of a X25 gene by measuring a length of said repeat, wherein n for normal individuals ranges from 1-22 and n for affected individuals is 120.

11. The method of claim 10, wherein said size of said repeat is measured by restriction endonuclease digestion of sample DNA and Southern Blot analysis.

12. The method of claim 10. wherein said size of said repeat is determined by pulsed field gel electrophoresis.

13. The method of claim 10, wherein SEQ ID NO 29 and SEQ ID NO 30 are used in said detecting step.

14. The method of claim 10, wherein SEQ ID NO 31 and SEQ ID NO 32 are used in said detecting steps.

15. A method for detecting a GAA polymorphism in a first intron of an X25 gene comprising the steps of performing a PCR assay to produce amplified prpducts of said first intron of said X25 gene and measuring the length of said amplified products.

16. The method of claim 15, wherein SEQ ID NO 29 and SEQ ID NO 30 are used in said PCR assay.

17. The method of claim 15, wherein SEQ ID NO 31 and SEQ ID NO 32 are used in said PCR assay.

18. A method of screening individuals for a mutation that leads to Friedreich's ataxia. comprising the steps of sequencing DNA from an individual. and comparing said sequence from said individual to SEQ ID NOS 1-12 to determine what differences there are between said sequence from said individual and SEQ ID NOS 1-12.

19. A method of treating Friedreich's ataxia, comprising the step of administering a pharmacologic dose of a protein having an amino acid sequence substantially similar to SEQ ID NO 4 to an individual.

20. A method of treating Friedreich's ataxia, comprising administration to an individual of a nucleic acid vector containing an X25 gene capable of expression.

21. As a composition of matter, the molecule having SEQ ID NO 1.

22. As a composition of matter, the molecule having SEQ ID NO 2.

23. As a composition of matter, the molecule having SEQ ID NO 3.

24. As a composition of matter, the molecule having SEQ ID NO 4.

25. As a composition of matter, the molecule having SEQ ID NO 5.

26. As a composition of matter, the molecule having SEQ ID NO 6.

27. As a composition of matter, the molecule having SEQ ID NO 7.

28. As a composition of matter, the molecule having SEQ ID NO 8.

29. As a composition of matter, the molecule having SEQ ID NO 9.

30. As a composition of matter, the molecule having SEQ ID NO 10.

31. As a composition of matter, the molecule having SEQ ID NO 11.

32. As a composition of matter, the molecule having SEQ ID NO 12.

33. As a composition of matter, the molecule having SEQ ID NO 13.

34. As a composition of matter, the molecule having SEQ ID NO 14.

35. As a composition of matter, the molecule having SEQ ID NO 15.

36. As a composition of matter, the molecule having SEQ ID NO 16.

37. As a composition of matter, the molecule having SEQ ID NO 17.

38. As a composition of matter, the molecule having SEQ ID NO 18.

39. As a composition of matter, the molecule having SEQ ID NO 19.

40. As a composition of matter, the molecule having SEQ ID NO 20.

41. As a composition of matter, the molecule having SEQ ID NO 21.

42. As a composition of matter, the molecule having SEQ ID NO 22.

43. As a composition of matter, the molecule having SEQ ID NO 23.

44. As a composition of matter, the molecule having SEQ ID NO 24.

45. As a composition of matter, the molecule having SEQ ID NO 25.

46. As a composition of matter, the molecule having SEQ ID NO 26.

47. As a composition of matter, the molecule having SEQ ID NO 27.

48. As a composition of matter, the molecule having SEQ ID NO 28.

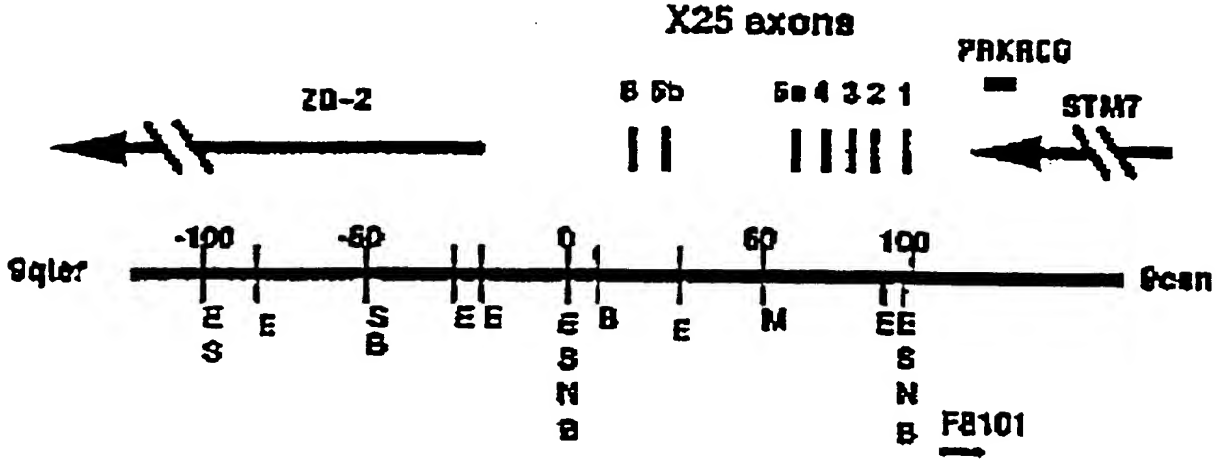
49. As a composition of matter. the molecule having SEQ ID NO 29.

50. As a composition of matter. the molecule having SEQ ID NO 30.

51. As a composition of matter, the molecule having SEQ ID NO 31.

52. As a composition of matter, the molecule having SEQ ID NO 32.

A



B

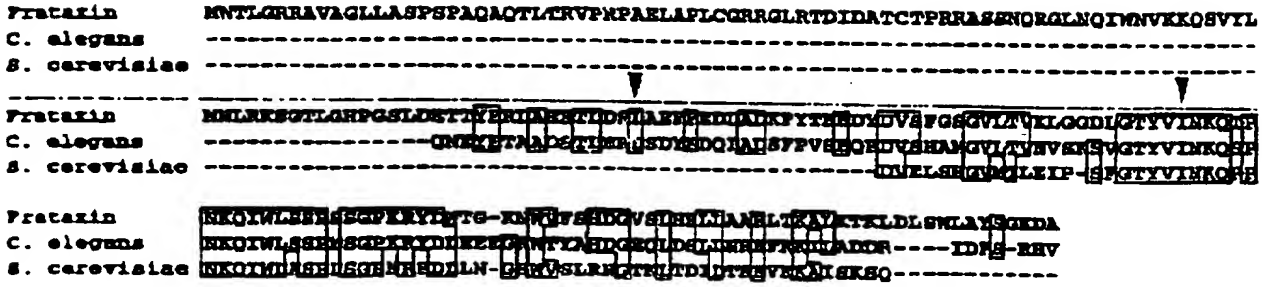


FIG. 1

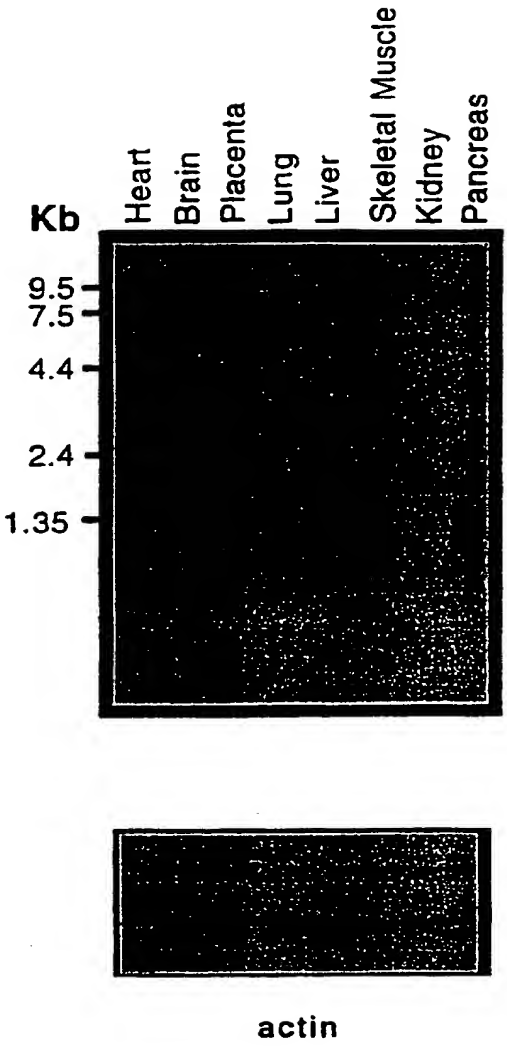


FIG. 2

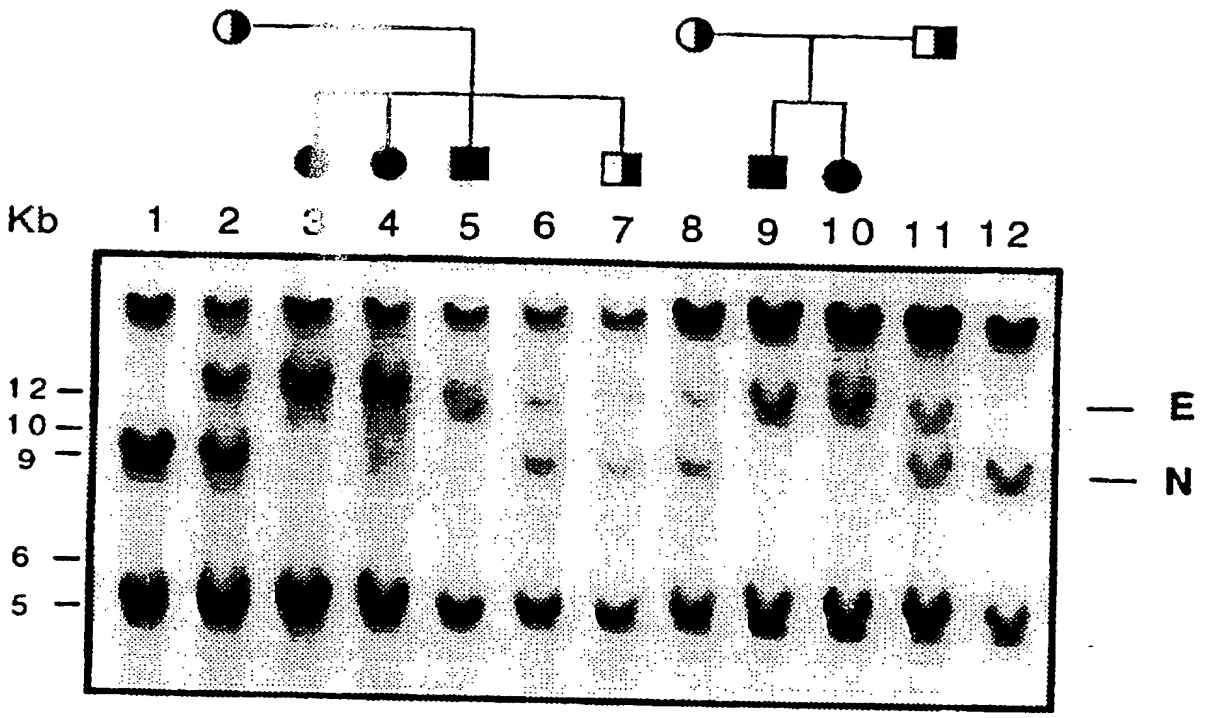


FIG.3

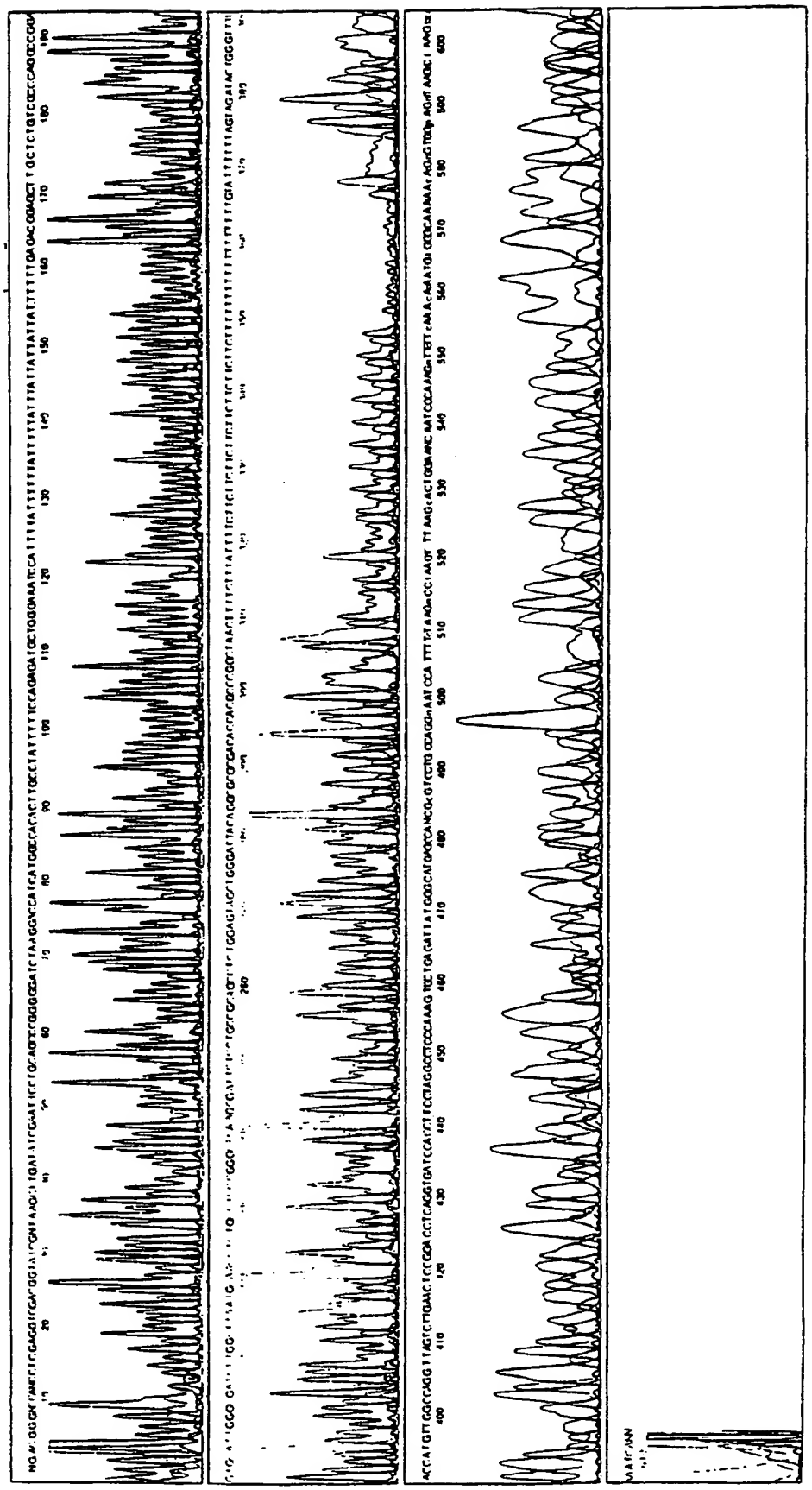


FIG.4

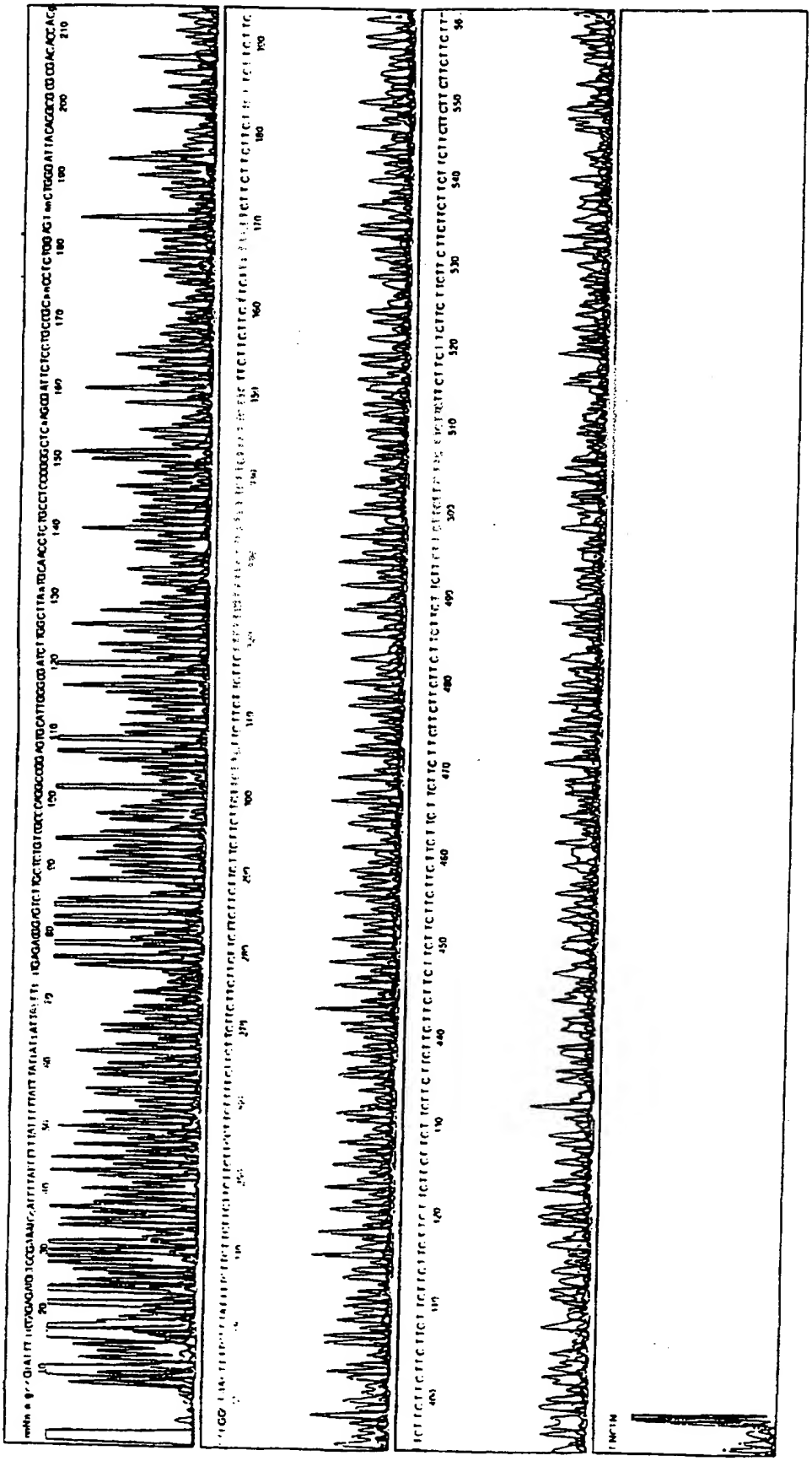


FIG. 5

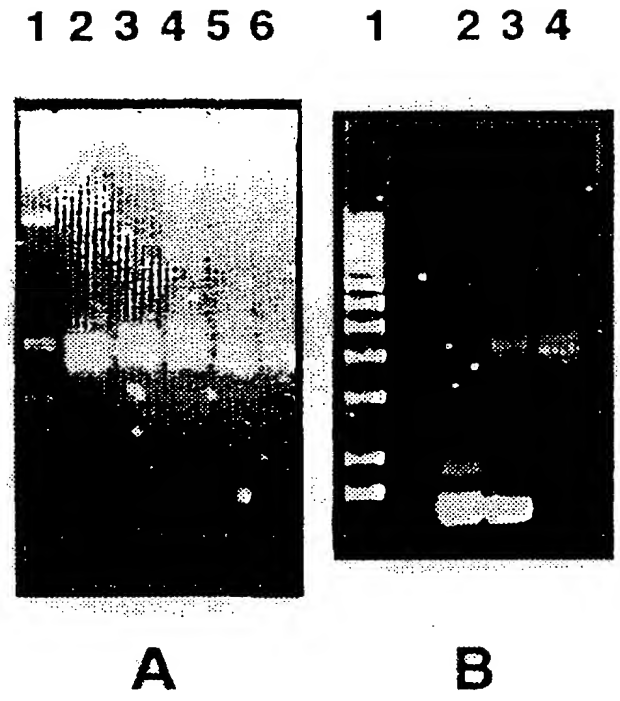


FIG. 6

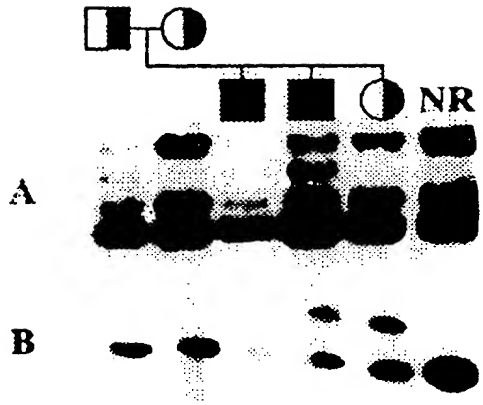


FIG. 7

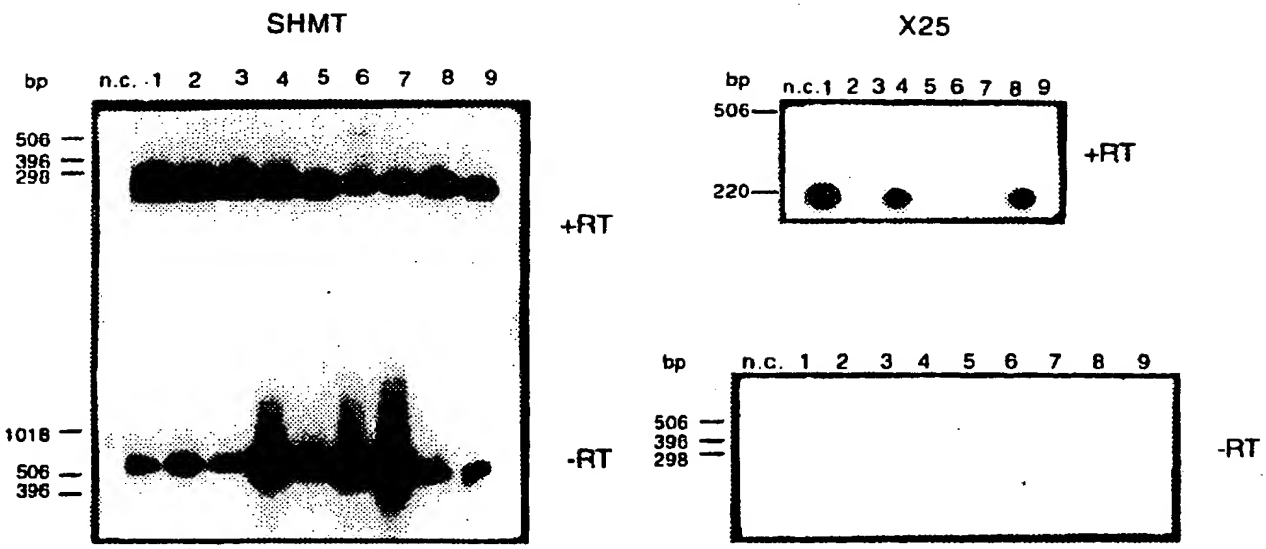


FIG. 8

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/EP 97/01070

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	AMERICAN JOURNAL OF HUMAN GENETICS, vol. 56, 1995, CHICAGO, US, pages 1116-1124, XP002035083 DOERFLINGER ET AL: "Ataxia with vitamin E deficiency" ---	
A	AMERICAN JOURNAL OF HUMAN GENETICS, vol. 57, 1995, CHICAGO, US, pages 1061-1067, XP002035084 MONTERMINI: "The Friedreich's Ataxia critical region spans a 150 Kb interval on chromosome 9q13" cited in the application --- -/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
11 July 1997	24.07.97

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Molina Galan, E

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Intern. ial Application No
PCT/EP 97/01070

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 90, 1993, WASHINGTON US, pages 109-113, XP002035085 DUCLOS ET AL: "Gene in the region of the Friedreich's Ataxia locus encodes a putative transmemebrane protein expressed in the nervous system" ---	
A	CELL, vol. 81, May 1995, NA US, pages 533-540, XP002035086 GACY ET AL: "Trinucleotide repeats that expand human disease form hairpin structures in vitro" cited in the application ---	
P,X	SCIENCE, vol. 271, 8 March 1996, LANCASTER, PA US, pages 1423-1427, XP002035087 CAMPUZANO ET AL: "Friedreich's ataxia: autosomal recesive disease caused by an intronic GAA triplet repeat expansion" see the whole document ---	1-52
P,X	NEW ENGLAND J MED, vol. 335, October 1996, pages 1169-1175, XP002035088 DÜRR ET AL: "Clinical and genetic abnormalities in patients with Friedreich's Ataxia" see the whole document ---	1-18
P,X	AMERICAN JOURNAL OF HUMAN GENETICS, vol. 59, September 1996, CHICAGO, US, pages 554-560, XP002035089 FILLA ET AL: "The relationship between GAA repeat length and clinical features in Friedreich's Ataxia" see the whole document ---	1-18
T	WO 97 05234 A (IMPERIAL COLLEGE ;CHAMBERLAIN SUSAN (GB); POOK MARK ADRIAN (GB); D) 13 February 1997 -----	

1

INTERNATIONAL SEARCH REPORT

In ational application No.
PCT/EP 97/ 01070

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Remark: Although claim(s) 19 and 20
is(are) directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern. al Application No
PCT/EP 97/01070

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9705234 A	13-02-97	NONE	